

Université de Montréal

**Segmentation automatique de la fibrose
pulmonaire sur images de tomodensitométrie en
radio-oncologie**

par

Nicolas Fréchette

Département de physique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Physique option physique médicale

Août 2019

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Segmentation automatique de la fibrose
pulmonaire sur images de tomodensitométrie en
radio-oncologie**

présenté par

Nicolas Fréchette

a été évalué par un jury composé des personnes suivantes :

Jean-François Arguin

(président-rapporteur)

Hugo Bouchard

(directeur de recherche)

Jacques A. de Guise

(codirecteur)

Stéphane Bedwani

(membre du jury)

(examineur externe)

Mémoire accepté le

6 novembre 2019

SOMMAIRE

La fibrose pulmonaire est une maladie pulmonaire interstitielle caractérisée par une production irréversible de tissus conjonctifs. Le pronostic de la maladie est plus faible que celui de plusieurs cancers. Dans les dernières années, cette pathologie a été identifiée comme un risque de complication suite à des traitements de radiothérapie. Développer une toxicité post-radique peut compromettre les bénéfices de la radiothérapie, ce qui fait de la fibrose pulmonaire une contre-indication relative. Localiser manuellement la présence de fibrose sur des images de tomodensitométrie (CT) est un problème difficile pouvant nécessiter l'intervention de plusieurs experts pour un seul patient.

L'objectif de ce projet est de segmenter automatiquement la fibrose pulmonaire sur des images CT. Des réseaux de neurones complètement convolutifs ont été développés et implémentés pour effectuer une assignation automatique de tissus pulmonaires. Sur une coupe axiale donnée en entrée, l'assignation est réalisée pour l'ensemble des voxels pulmonaires en une seule inférence. L'optimisation des paramètres a été réalisée dans des contextes d'apprentissage supervisé et semi-supervisé en minimisant des variantes de l'entropie croisée entre les prédictions et des annotations manuelles d'experts. Les données utilisées consistent en des images CT haute résolution ainsi que des délinéations réalisées par des radiologistes et des radio-oncologues. Les cartes de segmentation prédites ont été comparées par rapport à des segmentations manuelles afin de valider les tissus assignés par les réseaux convolutifs. Les résultats obtenus suggèrent que des applications en radio-oncologie sont envisageables, telles que le dépistage de la fibrose avant la planification de traitements et l'évaluation de la progression de la fibrose pendant et suivant les traitements de radiothérapie.

Mots-clés : Segmentation automatique, fibrose pulmonaire, réseaux de neurones, apprentissage machine

SUMMARY

Pulmonary fibrosis is an interstitial lung disease characterized by an irreversible production of scarring tissue. Pulmonary fibrosis has a particularly poor prognosis, with a mean survival after diagnosis lower than many cancers. This pathology was recently identified as a risk for complication following radiation therapy treatments. Pulmonary toxicity can lead to severe conditions that compromise the benefits provided by radiation therapy, making pulmonary fibrosis a relative contraindication to treatments. Manual segmentation of fibrosis on computed tomography (CT) images is a difficult task that can involve many experts for a single patient.

The aim of this project is to perform automatic segmentation of pulmonary fibrosis on CT images. Fully convolutional neural networks were developed and implemented to automatically assign lung tissues. For an input CT slice, every lung voxel is assigned a tissue in a single inference. Parameters optimization was performed in a supervised and semi-supervised manner by minimizing variants of the cross-entropy between the prediction and manual annotations produced by experts. The dataset employed consists of high resolution CT scans and delineations made by radiologists and radiation oncologists. Predicted segmentation maps were compared with manual segmentations to validate the tissues assigned by the convolutional networks. Results suggest that radiation oncology applications could be developed. Possible applications include pulmonary fibrosis screening prior to treatment planning and assessment of fibrosis progression during and post-treatment.

Keywords : Automatic segmentation, pulmonary fibrosis, neural networks, machine learning

TABLE DES MATIÈRES

Sommaire	iii
Summary	iv
Liste des figures	viii
Liste des tableaux	xi
Liste des sigles	xii
Liste des symboles	xiv
Remerciements	1
Chapitre 1. Introduction	2
1.1. Maladies pulmonaires interstitielles	3
1.2. Aperçu de la détection de maladies pulmonaires	5
1.3. Objectifs du projet	9
Chapitre 2. Tomodensitométrie	11
2.1. Interactions photon-matière	12
2.1.1. Diffusion Rayleigh	13
2.1.2. Effet photoélectrique	13
2.1.3. Diffusion Compton	14
2.1.4. Production de paires	15
2.2. Images de tomodensitométrie	15
2.2.1. Images haute résolution et de radio-oncologie	17
Chapitre 3. Réseaux de neurones	19
3.1. Neurone artificiel	19
3.2. Réseaux de neurones	23
3.2.1. Apprentissage	23
3.2.1.1. Optimisation d'une fonction coût	25
3.2.1.2. Descente de gradient stochastique avec momentum	26

3.2.1.3.	Estimation adaptative de moments	27
3.2.1.4.	Rétropropagation.....	28
3.2.1.5.	Fonctions d'activation	29
3.2.1.6.	Fonctions de coût.....	31
3.3.	Surapprentissage	32
3.3.1.	Arrêt prématuré.....	34
3.3.2.	Augmentation de données.....	34
3.3.3.	Dégradation des poids	36
3.3.4.	Abandon	36
3.4.	Réseaux de neurones à convolution	37
3.4.1.	Paramètres et activations.....	38
3.4.1.1.	Champ récepteur local et partage de paramètres.....	38
3.4.1.2.	Cartes d'activations	40
3.4.2.	Types d'opérations et de couches	42
3.4.2.1.	Convolution	42
3.4.2.2.	Pooling.....	46
3.4.2.3.	Couche connectée.....	47
3.4.2.4.	Normalisation de mini-lots.....	47
3.4.2.5.	Connexion raccourci.....	49
3.4.2.6.	Apprentissage par transfert.....	50
3.5.	Segmentation sémantique.....	52
3.5.1.	Architecture U-Net	55
Chapitre 4.	Méthodologie	58
4.1.	Images des patients et annotations.....	58
4.1.1.	Description des données requises.....	58
4.1.2.	Ensemble de données	60
4.2.	Évaluation des performances.....	63
4.2.1.	Matrice de confusion	63
4.2.2.	Métriques dérivées.....	64
4.3.	LungNet.....	66
4.3.1.	Architecture	66
4.3.2.	Détails d'implémentation	68

4.4. UNet-Progressive	68
4.4.1. Architecture	68
4.4.2. Entraînement	70
4.4.2.1. Entraînement avec régularisation	71
Chapitre 5. Résultats	73
5.1. Cartes de segmentation	74
5.2. Performances de la segmentation automatique	77
5.3. Évaluation de futures applications	83
Chapitre 6. Discussion	88
Chapitre 7. Conclusion	101
Bibliographie	103

LISTE DES FIGURES

1.1	Les diagnostics histologiques d'ILD les plus communs.....	3
1.2	Images CT pulmonaires contenant des changements radiologiques associés à la présence de maladies pulmonaires.....	4
1.3	Difficultés rencontrées dans l'interprétation d'images CT contenant des pathologies pulmonaires	5
2.1	Tomodensitomètre utilisé cliniquement.....	11
2.2	Importance relative des trois principales interactions de photons .	12
2.3	Impact du fenêtrage sur l'affichage d'une image CT	16
2.4	Images CT haute résolution et de radio-oncologie	17
3.1	Modèle du perceptron	20
3.2	Exemple de perceptron multicouche	24
3.3	Graphique des principales fonctions d'activation	30
3.4	Augmentation de données intégrée à l'entraînement.....	35
3.5	Utilisation de l'abandon neuronal	37
3.6	Organisation des neurones dans un réseau à convolution	39
3.7	Exemples de cartes d'activations.....	41
3.8	Insertion de zéros avant une convolution	42
3.9	Effet du pas de convolution sur le champ récepteur	43
3.10	Dilatation d'un noyau de convolution	45
3.11	Problème de maillage et choix de la dilatation	46
3.12	Normalisations de mini-lot et d'instance	48
3.13	Utilisation d'une connexion raccourci	49
3.14	Étapes d'un apprentissage par transfert.....	50
3.15	Différences entre la classification et la segmentation sémantique..	51
3.16	Opération de convolution transposée.....	53
3.17	Types d'annotations en classification et en segmentation	54

3.18	Architecture U-Net	56
3.19	Exemples de cartes d'activations d'un U-Net entraîné	57
4.1	Exemples d'annotations et d'images annotées	60
4.2	Masques provenant de la base de données	61
4.3	Images des patients de test	62
4.4	Matrice de confusion à 3 classes	64
4.5	Différents types d'assignations possibles	65
4.6	Réseau totalement convolutif LungNet	67
4.7	Réseau totalement convolutif UNet-Progressive	69
5.1	Carte de segmentation avec réalité de terrain saine	73
5.2	Carte de segmentation avec réalité de terrain de verre dépoli	75
5.3	Carte de segmentation avec réalité de terrain de fibrose	75
5.4	Segmentation d'une coupe avec motifs plus subtils	76
5.5	Coupe segmentée similairement par 3 CNN	77
5.6	Segmentation d'une coupe par 3 CNN avec fluctuations importantes	78
5.7	Cartes d'activations softmax d'une image avec annotations saines	78
5.8	Cartes d'activations softmax d'une image contenant de la fibrose	79
5.9	Cartes d'activations softmax d'une image contenant du verre dépoli	79
5.10	Matrices de confusion 3 classes calculées sur différents volets	81
5.11	Matrices de confusion de toutes les données de validation	82
5.12	Région pouvant bénéficier d'un seuillage softmax	83
5.13	Segmentations du premier patient de test	84
5.14	Segmentations du deuxième patient de test	85
5.15	Segmentations du troisième patient de test	87
6.1	Région avec vaisseaux correctement segmentés	88
6.2	Établissement d'une frontière entre les tissus sains et le verre dépoli	89
6.3	Distinctions plus subtiles faites par le réseau	90
6.4	Bordure de tissus sains assignée en tant que fibrose	91

6.5	Assignations erronées sur l'arbre trachéobronchique	92
6.6	Confusion entre le verre dépoli et la fibrose	93
6.7	Exemple de segmentation bruitée	93
6.8	Cas limites de verre dépoli	94
6.9	Régions ambiguës à proximité de l'arbre trachéobronchique	97

LISTE DES TABLEAUX

1.1	Comparaison des différentes méthodes d'assignation de tissus	9
2.1	Paramètres d'images CT haute résolution et de radio-oncologie ..	18
4.1	Textures pulmonaires incluses dans chacune des classes considérées pour la segmentation	59
4.2	Paramètres d'images et statistiques des données d'entraînement et de validation	62
4.3	Données de test disponibles pour les différents tissus	63
4.4	Dimensions des volumes d'activations du UNet-Progressive	71
5.1	Coefficients Dice et justesses d'assignation pour différents volets .	80
5.2	Performances des UNet-Progressive, LungNet et UNet-Progressive-Reg sur les données de validation	82
5.3	Activations softmax selon le type d'assignation	83
5.4	Performance du réseau sur les différents patients de test	86
5.5	Performances du réseau sur l'ensemble de test	86

LISTE DES SIGLES

Adam	Estimation adaptative de moments	<i>Adaptative moment estimation</i>
CHUM	Centre hospitalier de l'Université de Montréal	<i>University of Montreal Health Centre</i>
CNN	Réseau neuronal convolutif	<i>Convolutional neural network</i>
CT	Tomodensitométrie	<i>Computed tomography</i>
HRCT	Tomodensitométrie haute résolution	<i>High resolution computed tomography</i>
HUG	Hôpitaux Universitaires de Genève	<i>University Hospitals of Geneva</i>
ILD	Pneumopathie interstitielle	<i>Interstitial lung disease</i>
k-NN	Algorithme des k plus proches voisins	<i>k-nearest neighbours algorithm</i>
LReLU	Unité rectifiée linéaire avec fuite	<i>Leaky rectified linear unit</i>
MLP	Perceptron multi-couches	<i>Multilayer perceptron</i>
MSE	Erreur quadratique moyenne	<i>Mean squared error</i>
NIST	Institut national des normes et de la technologie	<i>National Institute of Standards and Technology</i>
PReLU	Unité rectifiée linéaire paramétrique	<i>Parametric rectified linear unit</i>
ReLU	Unité linéaire rectifiée	<i>Rectified linear unit</i>
SABR	Radiothérapie stéréotaxique ablative	<i>Stereotaxic ablative radiotherapy</i>
SGD	Descente de gradient stochastique	<i>Stochastic Gradient Descent</i>

SGDM	Descente de gradient stochastique avec momentum	<i>Stochastic gradient descent with momentum</i>
SVM	Machine à vecteurs de support	<i>Support vector machine</i>
UAT	Théorème d'approximation universelle	<i>Universal approximation theorem</i>

LISTE DES SYMBOLES

A	Justesse d'assignation d'un modèle
a	Activation d'un neurone
\mathcal{A}	Ensemble des voxels annotés
α	Paramètre de fuite
B	Paramètre de décalage
b	Biais d'un neurone
β_1	Facteur de décroissance du gradient
β_2	Facteur de décroissance du carré du gradient
CS	Carte de segmentation
D	Facteur de dilatation d'une convolution
δ_j^l	Erreur d'un neurone
DSC	Coefficient de Dice
ϵ	Constante de stabilité numérique
f	Coefficient de momentum
FN	Assignations de type faux négatif
FP	Assignations de type faux positif
Γ	Paramètre d'échelle
γ	Poids de classes
H_l	Largeur des cartes d'activations
HU	Unités d'Hounsfield
J	Fonction de coût supervisée

K	Dimension d'un filtre de convolution
L	Nombre de couches d'un réseau de neurones
l	Épaisseur d'un milieu homogène
λ	Facteur de régularisation
\mathcal{L}	Fonction de coût semi-supervisée
m	Taille des mini-lots d'entraînement
M	Matrice de confusion
μ	Coefficient d'atténuation linéaire
N_a^l	Nombre de cartes d'activations d'une couche de convolution
N_c	Nombres de classes distinctes
\mathcal{N}	Ensemble des voxels non-annotés
Ω	Pénalité sur la norme des poids
P	Épaisseur d'une bordure de zéros
Pr	Précision d'un modèle
Re	Recouvrement d'un modèle
s	Pas d'une convolution
S	Seuil d'un neurone perceptron
So	Activation softmax
θ	Paramètre de réseau
TP	Assignations de type vrai positif
\boldsymbol{v}	Vitesse dans l'espace des paramètres
W_l	Largeur des cartes d'activations d'une couche de convolution
w	Tenseur de poids
\boldsymbol{x}	Entrée d'un réseau de neurones

x	Entrées binaires d'un perceptron
χ	Paramètre de coût non supervisé
ξ	Fonction d'activation
\mathbf{y}	Vérité terrain d'une entrée
$\hat{\mathbf{y}}$	Segmentation d'une entrée par un réseau
Z	Numéro atomique
ζ	Sortie d'une couche de convolution
z_j^l	Entrée pondérée d'un neurone

REMERCIEMENTS

Je tiens tout d'abord à remercier Stéphane Bedwani pour m'avoir supervisé ainsi que pour son aide essentielle tout au long du projet. J'apprécie également le support apporté par mon directeur Hugo Bouchard et mon co-directeur Jacques de Guise, ainsi que d'avoir pu intégrer le Laboratoire d'orthopédie et d'imagerie du Centre de recherche du CHUM. Je remercie Dominique Mathieu, Houda Bahig, Marie-Pierre Campeau et Édith Fillion pour leur contribution essentielle en lien avec l'obtention et la manipulation des données des patients. Je tiens à souligner l'aide précieuse en matière d'intelligence artificielle apportée par Ismail Ben Ayed et Jose Dolz, professeurs à l'École de technologie supérieure. Je veux particulièrement souligner le soutien essentiel que m'ont apporté mes parents tout au long de mes études. Finalement, je remercie spécialement Chloé d'avoir été là à chacune des étapes et pour son support indispensable durant la totalité de mes études universitaires.

Chapitre 1

INTRODUCTION

Au Canada, 14 % des cancers sont des cancers pulmonaires, faisant de ce type de cancer le plus diagnostiqué après les cancers du sein et de la prostate. Le taux de survie des patients atteints d'un cancer pulmonaire est faible, soit de 17 % après 5 ans [1]. Dans le contexte du traitement des tumeurs pulmonaires par radiothérapie, livrer une dose de rayonnement plus élevée à la tumeur permettrait d'améliorer le taux de survie de façon importante [2]. Cependant, la dose maximale pouvant être prescrite à la tumeur est limitée par le fait que lors du traitement, les tissus externes à la tumeur sont également irradiés. L'irradiation de ces tissus implique un risque de développer d'importants effets secondaires pouvant mener à de sévères complications. Ces complications peuvent grandement diminuer les bénéfices thérapeutiques offerts par la radiothérapie. Un moyen entrepris pour limiter les risques de complication est de fixer des limites de dose aux différents organes à risque présents autour de la tumeur.

Aussi, en réalité, il est fréquent qu'un patient devant être traité pour un cancer pulmonaire soit atteint d'une pneumopathie interstitielle (ILD) touchant certaines régions des poumons. Les ILD sont des facteurs de risque de complications post-radiques, et ce même avec des traitements de radiothérapie stéréotaxique ablative (SABR), où les marges de planification sont grandement réduites [3]. Les risques de complication pourraient aussi être diminués en tenant compte de la localisation des pathologies pulmonaires existantes. Cependant, les types de tissus pulmonaires ne sont pas pris en considération dans la planification de traitement actuelle. Les traitements sont donc une contre-indication relative pour les patients atteints d'une ILD [4]. Tenir compte de la présence des ILD dans la planification de traitement est une étape importante pour diminuer les risques de complication et améliorer davantage le traitement des cancers du poumon par radiothérapie. Or, l'intégration de cette information dans le processus clinique nécessite de pouvoir détecter et localiser la présence de tissus pulmonaires pathologiques de façon efficace. Réaliser la

segmentation automatique de la fibrose pulmonaire sur des images de tomodensitométrie (CT) est l'objectif principal de ce projet.

1.1. MALADIES PULMONAIRES INTERSTITIELLES

La famille des ILD inclut plus de 300 types de maladies pouvant s'attaquer aux tissus interstitiels des poumons, c'est-à-dire les tissus qui entourent les alvéoles pulmonaires. Ces maladies peuvent, entre autres, mener à un excès de tissus conjonctifs dans les poumons. L'étendue des causes menant à une ILD est large, incluant notamment la prise de médicaments et les infections. La plupart des ILD font partie du groupe des pneumopathies interstitielles diffuses idiopathiques (*idiopathic interstitial pneumonia*), dont la cause est inconnue [5]. La place qu'occupent ces maladies idiopathiques dans la famille des ILD est présentée à la figure 1.1. Parmi les maladies de ce groupe, la plus commune est la fibrose pulmonaire.

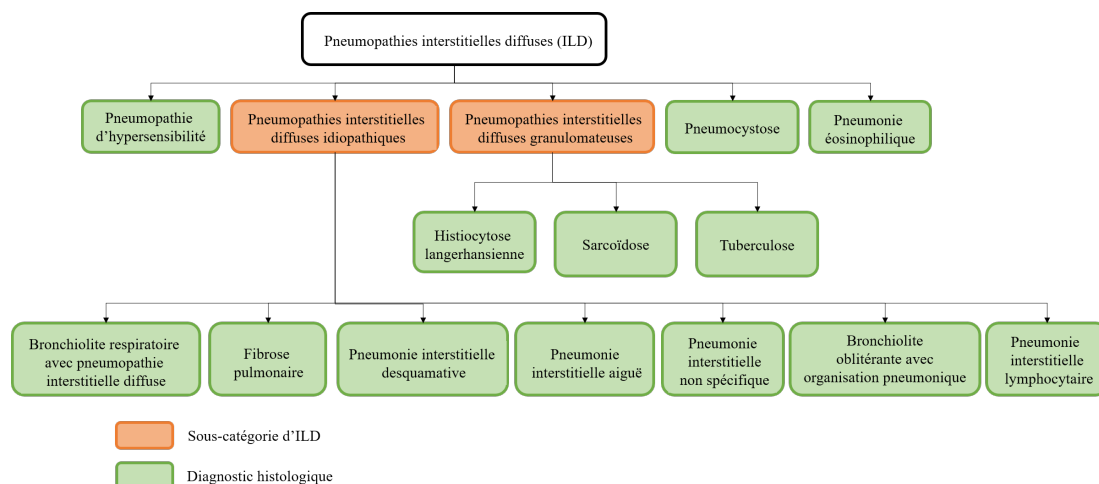


FIGURE 1.1. Les diagnostics histologiques d'ILD les plus communs. Figure tirée de la référence [6].

La fibrose pulmonaire est une maladie chronique qui implique une production excessive de tissus conjonctifs fibreux dans l'interstice pulmonaire. Le pronostic de la maladie est plus faible que celui de plusieurs cancers [7, 8]. La présence de ce type de tissus implique un raidissement, un épaissement et une cicatrisation des poumons qui engendrent une dégradation progressive des fonctions pulmonaires. Pour le diagnostic de la maladie, la méthode privilégiée

est l'interprétation d'images de tomodensitométrie haute résolution (HRCT) en raison de la nature non invasive de l'examen et l'excellente qualité des caractéristiques représentées sur les images.

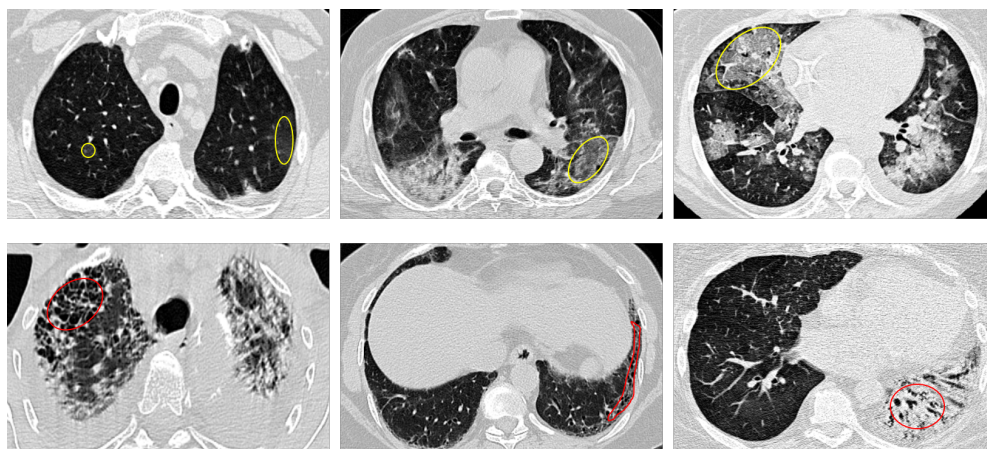


FIGURE 1.2. Images CT pulmonaires contenant des changements radiologiques associés à la présence de maladies pulmonaires. En haut, des images contenant des textures typiques de verre dépoli (contours en jaune) sont présentées. En bas, les contours en rouge présentent des motifs de rayons de miel (image de gauche), de la réticulation (au centre) et une présence intense de tissus conjonctifs (à droite), tous associés à de la fibrose.

Les maladies pulmonaires sont associées à des changements importants perceptibles sur les images radiologiques. En effet, ces changements sont visibles sur des images de tomodensitométrie sous forme de différents motifs (textures pulmonaires) qui sont parfois uniques à une maladie. Ces motifs peuvent alors être interprétés comme une signature de la maladie à laquelle ils sont associés. Par exemple, la fibrose pulmonaire est majoritairement caractérisée par des motifs de réticulation et de rayons de miel, mais aussi par de la bronchectasie et par des distorsions. Il est possible de caractériser l'ensemble de ces motifs par un seul type de tissu « fibrose », qui s'interprète comme un tissu associé à de la fibrose pulmonaire [6]. À l'inverse, des motifs non spécifiques tels que le verre dépoli sont perceptibles pour la majorité des ILD. Des images HRCT de patients présentant des tissus pathologiques sont illustrées à la figure 1.2. La complexité du problème d'assignation de tissus est notamment due à la grande diversité des textures pulmonaires de verre dépoli et de fibrose. Par exemple,

dans les stades peu sévères de verre dépoli, les tissus peuvent facilement être confondus avec des tissus sains. À l'inverse, les régions atteintes de verre dépoli sévère peuvent être confondues avec des tissus de fibrose. D'autres difficultés rencontrées lors de l'interprétation d'images CT contenant des pathologies pulmonaires sont présentées à la figure 1.3.

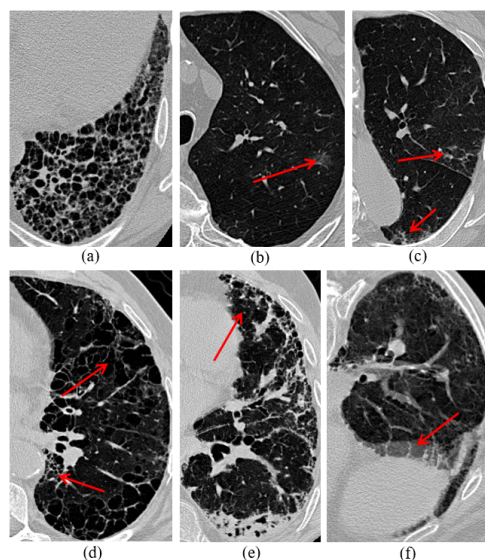


FIGURE 1.3. Bien que les motifs de fibrose pulmonaire puissent apparaître comme des (a) rayons de miel facilement distinguables, des poumons pathologiques contiennent souvent des motifs (b) subtils, (c) petits et (d) ambigus. (e) Un poumon qui contient une étendue moyenne de fibrose peut mener à des régions saines incertaines. (f) Des artéfacts de mouvement, fréquemment rencontrés en imagerie pulmonaire, peuvent introduire des difficultés supplémentaires dans l'assignation des tissus.

1.2. APERÇU DE LA DÉTECTION DE MALADIES PULMONAIRES

Actuellement, l'interprétation et l'annotation de pathologies pulmonaires sont majoritairement effectuées par inspection visuelle. En radio-oncologie, aucun algorithme permettant la segmentation automatique de maladies pulmonaires n'est disponible sur les outils commerciaux de visualisation et de délimitation de régions anatomiques. L'analyse des différents motifs et l'assignation des types de tissus devraient donc être effectuées manuellement. Cependant, pour un seul patient, la quantité d'images à analyser peut être très importante, de l'ordre de plusieurs centaines d'images. Le volume de données à analyser et la difficulté du

problème en général font en sorte qu’une segmentation manuelle peut exiger plusieurs heures de segmentation et plusieurs experts pour un seul patient [6]. Pour ces raisons, en radio-oncologie, les motifs de fibrose pulmonaire ne sont pas segmentés. En raison des difficultés liées à l’assignation de tissus pulmonaires pathologiques, des approches sont développées dans le but de réduire l’effort manuel requis. Les prochains paragraphes visent à faire une revue des principales approches visant à faciliter la détection des maladies pulmonaires interstitielles. Le tableau 1.1 compare sommairement les différentes méthodes discutées ci-dessous dans le but de faire ressortir les principaux avantages et les contraintes de chacune des méthodes.

D’abord, des études se basent sur l’analyse des premiers moments statistiques de l’histogramme des nombres CT dans la région pulmonaire. Celles-ci visent à établir les différences entre l’histogramme des nombres CT de patients sains avec celui de patients atteints d’ILD. Par exemple, des études ont établi que le premier moment de l’histogramme des nombres CT de la région pulmonaire est plus élevé pour des patients atteints de sarcoïdose pulmonaire que pour des patients sains [9]. Une étude similaire a démontré que pour des patients atteints de fibrose pulmonaire idiopathique, le kurtosis et l’asymétrie de l’histogramme sont moins élevés [10]. De telles études permettent de fournir une information pertinente à l’expert afin de l’aider à faire l’interprétation des images du patient. D’autres approches se basent sur des informations fournies *a priori*, telles que des seuils. Les approches basées sur les seuils consistent à assigner les tissus aux voxels selon leur intensité. Cette approche est la plus utilisée pour quantifier l’emphysème auprès de patients atteints de maladies pulmonaires obstructives chroniques [11]. Par exemple, pour cette application particulière, il a été déterminé qu’un seuil de -950 HU pouvait être utilisé pour quantifier l’emphysème de façon satisfaisante [12, 13]. Dans le même ordre d’idées, des changements de fibrose liés à une pneumonie, la fibrose pulmonaire interstitielle et les motifs en rayons de miel peuvent être quantifiés à l’aide de seuils basés sur plusieurs valeurs [14–16]. Selon le type de la maladie sous-jacente, divers seuils ont également été établis pour la quantification du verre dépoli [11]. En résumé, les méthodes basées sur les seuils permettent d’avoir un aperçu de la présence de certains types de tissus associés à des maladies pulmonaires. Cependant, ces méthodes offrent un résultat très approximatif et, en plus d’être difficiles à

déterminer de façon adéquate, les seuils sont spécifiques à une condition ou à une maladie particulière. Celles-ci contribuent à faciliter la détection de la fibrose comparativement à une segmentation manuelle. Par contre, une implication importante de la part des médecins est toujours nécessaire et l'information fournie n'est que très approximative.

La détection des ILD a également été approchée par des méthodes d'analyse statistique de textures. Sur une région d'intérêt donnée, une texture correspond à une mesure quantitative de l'arrangement des niveaux de gris. Ces approches impliquent de déterminer manuellement un ensemble de textures qui convient pour le problème donné. Pour donner quelques exemples, des textures peuvent être extraites de l'histogramme des niveaux de gris. Cette catégorie de textures inclut généralement les premiers moments statistiques de l'histogramme. D'autres textures peuvent être calculées à partir d'une matrice de co-occurrence, c'est-à-dire une matrice qui indique le nombre de fois qu'une paire de voxels est répétée dans la région. À partir de cette matrice, des textures telles que l'énergie, l'entropie et le contraste de l'image peuvent être extraites. D'autres textures importantes, telles que la non-uniformité des longueurs de segment, sont obtenues à l'aide d'une matrice de longueur de segment, qui présente l'occurrence d'une séquence continue de voxels de même intensité selon une direction donnée. Une fois un ensemble de textures déterminé, celles-ci sont calculées sur une région d'intérêt extraite d'une image et utilisées pour déterminer le type de tissu représenté par la région. Généralement, ces textures sont envoyées en entrée à un modèle d'apprentissage machine qui effectue une classification de la région. Des modèles utilisés dans la classification de maladies pulmonaires incluent les machines à vecteurs de support (SVM), les classificateurs bayésiens et l'algorithme des k plus proches voisins (k -NN) [17–21]. Des perceptrons multicouches (réseaux de neurones) ont également été utilisés comme classificateurs de textures [22–25]. L'analyse de textures est plus performante que les approches basées sur les seuils, pouvant offrir des performances qui s'approchent de celles obtenues par l'évaluation visuelle d'un radiologiste [11]. De plus, une fois qu'un bon ensemble de textures à extraire est déterminé, il y a relativement peu d'efforts manuels à fournir par les médecins. Aussi, comme l'analyse ne repose plus sur une interprétation visuelle des images, les problèmes liés aux variations dans les interprétations

réalisées par différents observateurs sont moins présents. Toutefois, ces méthodes ne sont pas complètement automatiques. Aussi, l’assignation de tissus est effectuée sur des régions d’intérêt suffisamment grandes pour extraire des textures cohérentes et non au niveau des voxels. Déterminer un ensemble satisfaisant de textures statistiques est relativement complexe, puisqu’un nombre important de textures peuvent être considérées, plusieurs sont propres au problème étudié et il est généralement possible d’extraire les textures selon des directions arbitraires.

D’autres approches abordent l’assignation des types de tissus pulmonaires par l’apprentissage profond. En apprentissage profond, l’effort manuel à fournir est grandement réduit au point où l’analyse peut être considérée comme étant automatique. En effet, de tels modèles apprennent les représentations des données qui permettent de résoudre efficacement le problème considéré. En d’autres mots, les textures (ou caractéristiques) adéquates sont déterminées par le modèle plutôt que par l’expert. L’apprentissage profond est discuté de façon plus détaillée au cours du chapitre 3. Une première approche consiste à extraire des régions d’intérêt sur les poumons et à fournir les nombres CT de la région à un réseau de neurones à convolution (CNN) [26, 27]. Tout comme avec l’analyse de textures traditionnelle, ces méthodes permettent d’obtenir de l’information sémantique pertinente sur la région. En effet, la sortie d’un CNN nous informe globalement sur le type de motif qui peut être associé à la région. Toutefois, la classification de régions d’intérêt par un CNN se fait toujours au niveau de régions d’intérêt et non au niveau des voxels.

Plus récemment en apprentissage profond, des réseaux de neurones complètement convolutifs ont été développés pour faire la segmentation sémantique d’images. Ces réseaux peuvent être utilisés pour faire des prédictions denses (au niveau des voxels) de façon automatique et en une seule inférence. Appliqués aux ILD, ces réseaux permettent donc d’assigner directement un tissu à chacun des voxels des poumons. Ces segmentations sont très fines, puis fournissent de l’information sur les types de tissus présents (information sémantique) et la localisation des différents motifs (information spatiale). Les réseaux de neurones complètement convolutifs sont particulièrement intéressants pour des applications en interprétation d’images

médicales. Le potentiel d'un réseau utilisé en segmentation sémantique a été évalué pour la segmentation de certaines pathologies pulmonaires, telles que les micronodules [28]. Les réseaux complètement convolutifs, utilisés au cours de projet, sont abordés dans le chapitre 3.

TABLEAU 1.1. Comparaison des principaux avantages et des contraintes qui distinguent les différentes méthodes d'assignation de tissus pulmonaires

Méthode	Avantages	Contraintes
Application de seuils	Rapide, facilité d'implantation	Aperçu approximatif
Class. de régions par analyse de textures	Semi-automatique, performant	Régions, textures manuelles
Class. de régions par CNN	Automatique, performant	Régions, données nécessaires
Segmentation sémantique par CNN	Assignation à des voxels	Données nécessaires

1.3. OBJECTIFS DU PROJET

L'objectif principal de ce projet est de segmenter automatiquement la fibrose pulmonaire sur des images CT pour de futures applications en radio-oncologie. Pour ce faire, la première partie du projet a été de réaliser une revue des réseaux de neurones, puis d'établir un ensemble de données à utiliser dans l'optimisation des paramètres des modèles et pour l'évaluation des performances.

Ensuite, des réseaux de neurones complètement convolutifs ont été développés et implémentés dans des contextes d'apprentissage supervisé et semi-supervisé pour générer des cartes de segmentation multiclassées de la fibrose pulmonaire. De plus, des données de test ont été produites au Centre hospitalier de l'Université de Montréal (CHUM) pour évaluer la possibilité de réaliser de futures applications en radio-oncologie basées sur une segmentation automatique de la fibrose.

L'apport principal de ce projet est le développement et l'implémentation de modèles pouvant réaliser la segmentation des tissus associés à de la fibrose pulmonaire de façon automatique. La présence et la localisation de tissus pathologiques dans les poumons ne sont pas prises en compte dans la planification des traitements de radiothérapie. Donc, cette étude est un pas vers l'intégration de cette information en radio-oncologie au CHUM.

Un outil de segmentation automatique ouvre également la porte à de nombreuses applications cliniques, telles que le dépistage de la fibrose. Le projet a également mené au développement d'une base de données de radio-oncologie qui pourra être bonifiée et utilisée dans de futurs travaux.

Dans ce mémoire, le chapitre 2 aborde brièvement la tomodesitométrie. Dans ce chapitre, les bases physiques et les paramètres des images sont discutés, afin de présenter les principales différences entre les images HRCT et les images de radio-oncologie. Le chapitre 3 porte sur l'apprentissage machine et, plus particulièrement, les réseaux de neurones. Les objectifs de ce dernier sont de présenter les concepts centraux à l'apprentissage profond et à la segmentation sémantique ainsi que de discuter plus en détail des techniques utilisées au cours du projet. Au cours du chapitre 4, la méthodologie utilisée est définie. Dans un premier temps, les données utilisées et produites sont abordées. Ensuite, les métriques utilisées, les architectures implémentées et des détails sur l'optimisation des paramètres sont discutées. Le chapitre 5 présente les différents résultats obtenus. Les résultats incluent les segmentations produites, une évaluation des performances et des comparaisons entre les différents modèles considérés. Finalement, les différents résultats et les applications envisageables en radio-oncologie sont discutés au chapitre 6, suivis d'une conclusion au chapitre 7.

Chapitre 2

TOMODENSITOMÉTRIE

La tomodensitométrie est une modalité d'imagerie à transmission de rayons X qui est couramment utilisée dans le domaine médical. Les images générées offrent un bon contraste entre les différents tissus et structures, ce qui en fait une modalité de choix pour la détection des motifs liés à la fibrose pulmonaire. La figure 2.1 présente un exemple de tomodensitomètre utilisé cliniquement. Dans ce premier chapitre théorique, la tomodensitométrie est brièvement abordée. Dans un premier temps, la modalité d'imagerie dans son ensemble et les principes physiques liés à l'acquisition des données sont discutés. Ensuite, quelques paramètres des images CT sont abordés dans l'optique de comparer les images HRCT avec celles de radio-oncologie.



FIGURE 2.1. Exemple d'un tomodensitomètre utilisé cliniquement. Image tirée de www.siemens-healthineers.com.

2.1. INTERACTIONS PHOTON-MATIÈRE

En résumé, la tomодensitométrie consiste à faire l'acquisition des données projectives à plusieurs angles autour d'un volume d'intérêt. Les données projectives sont obtenues à l'aide d'un tube à rayons X et de détecteurs situés dans l'anneau au centre duquel se trouve le patient. D'abord, le tube à rayons X émet des photons par bremsstrahlung. Lors du passage du faisceau dans le patient, une fraction des photons interagit avec les atomes du patient. Les photons transmis sont alors mesurés par les détecteurs. Pour une source monoénergétique et N_0 photons à l'entrée d'un milieu, le rayonnement primaire transmis N est donné par la loi d'atténuation :

$$N = N_0 e^{-\sum_i \mu_i l_i}, \quad (2.1.1)$$

où μ_i et l_i sont respectivement le coefficient d'atténuation linéaire et l'épaisseur du milieu i . Le coefficient d'atténuation linéaire, défini pour un milieu homogène, correspond à la probabilité par unité de distance qu'un photon interagisse avec un site d'interaction. Celui-ci dépend de l'énergie des photons ainsi que de la densité du milieu traversé. Les interactions possibles sont la diffusion Rayleigh (μ_{Ry}), l'effet photoélectrique (μ_{ph}), la diffusion Compton (μ_{C}) et la production de paires (μ_{pp}) [29]. En termes des coefficients d'atténuation linéaire propres à ces différentes interactions de photons, μ peut être exprimé ainsi [30] :

$$\mu = \mu_{\text{ph}} + \mu_{\text{Ry}} + \mu_{\text{C}} + \mu_{\text{pp}}. \quad (2.1.2)$$

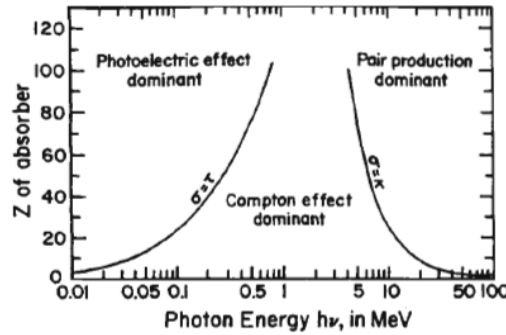


FIGURE 2.2. Importance relative des interactions photoélectrique (τ) et Compton (σ) ainsi que de la production de paires (κ). Les courbes indiquent les valeurs de $h\nu$ et de Z pour lesquelles deux types d'interactions sont de même importance. Graphique tiré de la référence [31].

Avec les énergies utilisées en tomодensitométrie et les numéros atomiques effectifs Z des tissus humains ($Z_{\text{tissus mous}} = 7$, $Z_{\text{os}} = 13$), l'interaction prédominante est l'effet Compton (figure 2.2) [31, 32]. Dans les prochaines sous-sections, les différentes interactions de photons sont brièvement détaillées.

2.1.1. Diffusion Rayleigh

La diffusion Rayleigh est caractérisée par l'interaction d'un photon avec le nuage électronique d'un atome, sans que l'atome subisse d'excitation ou d'ionisation. Le photon diffusé a donc la même énergie que le photon incident. Pour l'effet Rayleigh, la relation de proportionnalité donnée à l'équation 2.1.3 s'applique à la section efficace atomique $\sigma_{\text{a,Ry}}$:

$$\sigma_{\text{a,Ry}} \propto \frac{Z^2}{(h\nu)^2}. \quad (2.1.3)$$

Puis, pour le coefficient d'atténuation de l'interaction Rayleigh normalisé par la densité $\frac{\mu_{\text{Ry}}}{\rho}$ [33] :

$$\frac{\mu_{\text{Ry}}}{\rho} \propto \frac{Z}{(h\nu)^2}. \quad (2.1.4)$$

2.1.2. Effet photoélectrique

L'effet photoélectrique est une interaction qui se produit avec un électron des couches électroniques plus basses lorsqu'un photon est incident sur un atome. Le photon est complètement absorbé et cause un léger recul de l'atome ainsi que l'éjection d'un photoélectron pouvant par la suite exciter et ioniser le milieu. L'énergie cinétique du photoélectron est approximativement donnée par la différence entre l'énergie du photon incident et l'énergie de liaison de la couche électronique où se trouvait l'électron. Une particularité de l'effet est que les courbes de μ en fonction de $h\nu$ contiennent des discontinuités qui surviennent lorsque l'énergie du photon est légèrement supérieure à l'énergie de liaison d'une couche. Aux énergies où se trouvent ces discontinuités, la probabilité d'interaction devient alors très élevée. Ces discontinuités sont particulièrement importantes dans les éléments à haut Z .

En tomodesitométrie, des acquisitions peuvent être faites avec l'injection d'un agent de contraste à haut Z , tel que l'iode, pour profiter des raies d'absorption présentes dans les énergies effectives du CT. Cela permet d'avoir un meilleur contraste sur les images produites. Hors des discontinuités et de la région relativiste, la relation de proportionnalité donnée à l'équation 2.1.5 s'applique à $\sigma_{a,ph}$:

$$\sigma_{a,ph} \propto \frac{Z^5}{(h\nu)^{\frac{7}{2}}}. \quad (2.1.5)$$

De façon similaire, pour le coefficient d'atténuation de l'effet photoélectrique normalisé par la densité $\frac{\mu_{ph}}{\rho}$ [32] :

$$\frac{\mu_{ph}}{\rho} \propto \frac{Z^4}{(h\nu)^{\frac{7}{2}}}. \quad (2.1.6)$$

2.1.3. Diffusion Compton

La diffusion Compton se produit lorsqu'un photon interagit avec un électron du nuage électronique. Pour un photon incident d'énergie $h\nu$, le photon diffusé est émis à un angle θ et possède une énergie $h\nu'$, donnée par l'équation suivante :

$$h\nu' = \frac{h\nu}{1 + \frac{h\nu}{m_e c^2} (1 - \cos(\theta))}. \quad (2.1.7)$$

Lors de l'interaction, un électron de recul est éjecté avec un angle et une énergie cinétique qui dépendent de l'angle de diffusion et de l'énergie du photon incident. Comme une interaction Compton résulte en l'émission d'un photon à un angle différent de l'angle incident, cette interaction est nuisible en imagerie médicale. En effet, si ce rayonnement secondaire atteint le détecteur, l'image sera davantage bruitée. Celui-ci peut également être la cause de certains artéfacts en tomodesitométrie. Il est donc préférable d'utiliser des énergies pour lesquelles l'effet photoélectrique est dominant, tout en ayant une quantité suffisante de photons transmis pour avoir une qualité d'image adéquate. Aussi, une grille anti-diffusion peut être placée avant le détecteur afin de faire interagir une part importante des photons secondaires avec des lames de plomb. Ainsi, cette grille permet de réduire grandement la quantité de photons secondaires qui se rendent au détecteur.

2.1.4. Production de paires

Lorsque l'énergie d'un photon incident est supérieure à l'énergie de masse d'un électron et d'un positron, celui-ci peut être absorbé par le site d'interaction et une production de paires peut survenir. La production de paires consiste en l'émission d'un électron et d'un positron dans le milieu. Les deux particules émises peuvent alors déposer leur énergie jusqu'à la fin de leur parcours. La collision du positron avec un électron peut ensuite causer une annihilation et l'émission de deux photons de 511 keV approximativement antiparallèles. L'annihilation est centrale à la tomographie par émission de positrons, une modalité d'imagerie fonctionnelle utilisée en médecine nucléaire et en radio-oncologie. La production de paires intervient en radiothérapie, mais pas en imagerie par rayons X, étant donné que les photons utilisés sont nettement moins énergétiques que l'énergie minimale requise de 1.022 MeV pour la production de paire dans le champ du noyau et de 2.044 MeV dans le champ d'un électron (triplet).

2.2. IMAGES DE TOMODENSITOMÉTRIE

Lors de l'acquisition des projections, une rotation du tube et des détecteurs autour du patient et un déplacement de table sont effectués afin d'acquérir des projections à différents angles sur un volume d'intérêt. Ces projections sont ensuite utilisées pour reconstruire un volume pouvant être visualisé sous forme de coupes axiales. De nombreuses méthodes existent pour effectuer la reconstruction. Celles-ci sont généralement basées sur le théorème de la tranche centrale de Fourier, sur la rétroprojection filtrée ou sur les techniques de reconstruction itérative [34]. En résumé, chaque projection acquise lors de l'examen est associée à un angle et à un élément de détection, puis toutes les projections peuvent être représentées par une matrice appelée sinogramme. Le sinogramme peut ensuite servir de point de départ pour la reconstruction, où chacune de ses lignes pourrait, par exemple, être convoluée avec un filtre dans le domaine spatial et rétroprojetée sur l'image en train d'être reconstruite.

De plus, sur les images de tomodensitométrie, les nombres CT attribués aux voxels ont une signification physique. Les nombres CT sont exprimés en unités d'Hounsfield (HU) et sont reliés au coefficient d'atténuation linéaire par l'équation 2.2.1. De plus, étant donné la relation entre le coefficient d'atténuation et la densité du milieu, il est possible de convertir les nombres CT en densités électroniques, qui sont utilisées dans la planification de traitement.

$$\text{Nombre CT} = \frac{1000 (\mu - \mu_{\text{eau}})}{\mu_{\text{eau}} - \mu_{\text{air}}} \text{HU}. \quad (2.2.1)$$

Un voxel contenant de l'air aura une valeur près de 1000 HU et sera noir sur l'image. Inversement, les os, qui ont un Z et un μ plus élevés, sont beaucoup plus clairs (blancs). Les tissus mous se retrouvent entre ces deux extrêmes, sous différentes teintes de gris. La plage complète des 4096 valeurs de HU est trop riche pour être exprimée en niveaux de gris distinguables visuellement. Cela peut être vu comme une limitation des experts qui analysent les images, qui n'est pas directement présente pour les algorithmes d'apprentissage machine. En général, l'affichage des images se fait en fixant une fenêtre définie par un centre et une largeur. La figure 2.3 illustre un exemple de l'impact du fenêtrage utilisé.

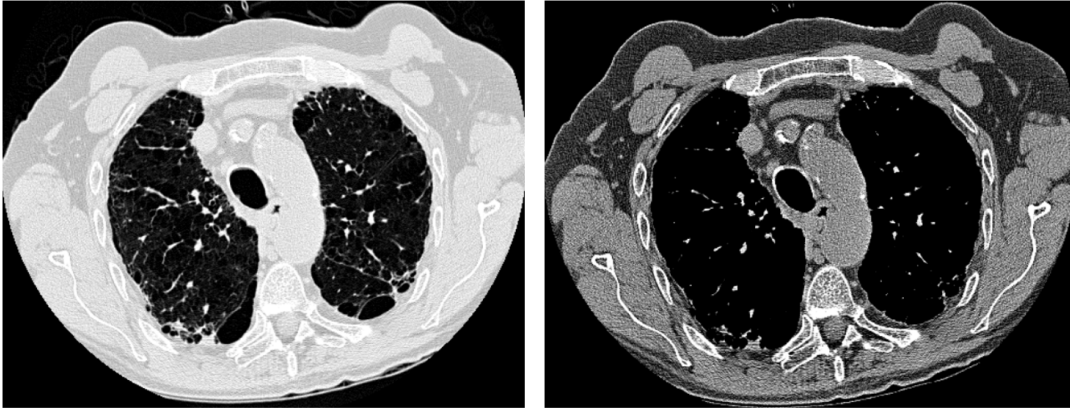


FIGURE 2.3. À gauche, une coupe axiale de poumons est affichée avec un fenêtrage pulmonaire (centre de -400 HU, largeur de 1200 HU). À droite, la même image est affichée avec un fenêtrage de tissus mous (centre de 40 HU, largeur de 400 HU).

2.2.1. Images haute résolution et de radio-oncologie

La qualité des images produites dépend de plusieurs paramètres. Parmi ceux-ci, notons le potentiel du tube à rayons X, le temps d'exposition, le courant du tube, le champ de vue, la dimension physique des voxels et l'épaisseur des coupes.

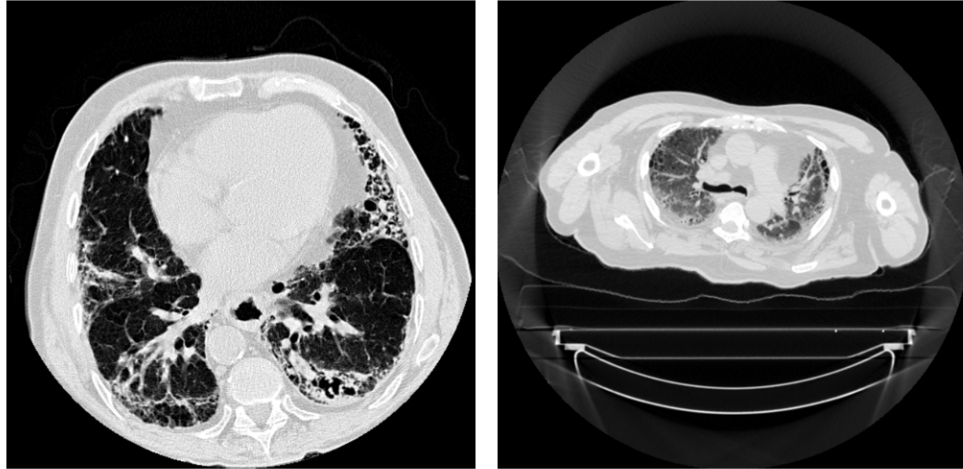


FIGURE 2.4. Images de patients atteints de fibrose pulmonaire provenant de bases de données haute résolution (à gauche, champ de vue de 336 mm) et de radio-oncologie (à droite, champ de vue de 600 mm). Un centre de -400 HU et une largeur de 1200 HU sont utilisés pour les deux fenêtrages.

Le choix de ces paramètres a des effets importants sur les images obtenues. Par exemple, une augmentation du potentiel du tube à rayons X a pour effet de diminuer le contraste généré sur les images. Également, au niveau du produit courant-temps, une augmentation de ce paramètre a pour effet de diminuer le bruit, étant donné que le nombre de photons produits et atteignant le détecteur est augmenté. Plusieurs paramètres des appareils et des images peuvent varier de façon considérable entre des images haute résolution et de radio-oncologie. De plus, pour obtenir des images de meilleure résolution, la tomодensitométrie haute résolution utilise une collimation étroite et des algorithmes de reconstruction à haute fréquence spatiale [35].

Ces différences ont un impact sur les coupes reconstruites, tel que présenté à la figure 2.4. Des paramètres d’images provenant de bases de données haute résolution et de radio-oncologie sont comparés au tableau 2.1. Les images CT haute résolution pulmonaires sont généralement caractérisées par un petit champ de vue et une petite taille de voxels. Aussi, dans le but d’optimiser la qualité d’image et la représentation des détails, ces paramètres peuvent être variés de façon importante d’un patient à l’autre. En radio-oncologie, le champ de vue est plutôt fixé à 500 mm dans la majorité des cas, pour la planification des traitements. Aussi, les coupes axiales de radiologie sont généralement plus minces.

TABLEAU 2.1. Champs de vue (FOV), potentiels de tube (kVp), produits courant-temps, épaisseur de coupes et tailles de voxels pour des images CT haute résolution (HRCT) et de radio-oncologie (R-O)

Paramètre	HRCT	R-O
FOV (mm)	278–499	500–700
kVp (kV)	120–140	120–140
Courant-temps (mAs)	72–264	80–1203
Épaisseur (mm)	1	2–3
Taille de voxel (mm)	0.42–0.97	0.98–1.37

Chapitre 3

RÉSEAUX DE NEURONES

La segmentation d'images, de même que les problèmes de classification, peuvent être approchés par l'apprentissage machine. La distinction entre ces deux types de problèmes est que la classification vise à catégoriser une donnée d'entrée par une seule classe alors que, dans les problèmes de segmentation, l'objectif est d'effectuer une prédiction par voxel. Pour une image de fibrose, une classification correspond à caractériser globalement l'image comme contenant des poumons sains ou atteints de fibrose, alors qu'avec une segmentation, chaque voxel se voit assigné un tissu sain ou de fibrose. Actuellement, les méthodes parmi les plus performantes qui approchent ces problèmes utilisent des architectures de réseaux de neurones profonds combinées à des algorithmes d'apprentissage. Malgré la différence considérable entre la segmentation et la classification, les réseaux de neurones utilisés dans les deux cas sont similaires. En fait, les réseaux utilisés dans les problèmes de segmentation sont généralement basés sur une adaptation des architectures utilisées pour la classification. Pour ces raisons, ce second chapitre théorique aborde les réseaux de neurones et l'apprentissage. Dans un premier temps, les neurones artificiels et les réseaux de neurones connectés, aussi appelés perceptrons multicouches, sont introduits. Ensuite, l'entraînement par apprentissage supervisé est abordé. Enfin, la deuxième partie de ce chapitre discute des réseaux de neurones à convolution utilisés en imagerie ainsi que de la segmentation sémantique.

3.1. NEURONE ARTIFICIEL

D'abord, un neurone artificiel constitue l'élément de base du réseau de neurones. Le premier modèle de neurone, appelé perceptron, a été développé par Rosenblatt en 1953, qui s'est inspiré des travaux de McCulloch et Pitts [36, 37]. L'idée fondamentale du perceptron est inspirée des neurones biologiques. Une des similitudes entre les perceptrons et les neurones biologiques est que les deux ont un comportement binaire. Lorsque les stimulus d'entrée sont faibles, le neurone demeure au repos et il s'active lorsque l'intensité dépasse un seuil.

Le perceptron, présenté à la figure 3.1, reçoit une série d'entrées binaires $x_{b,i}$, c'est-à-dire qu'elles prennent la valeur 0 ou 1. Ces entrées passent par des connexions auxquelles sont attribués des poids w_i . Les poids sont utilisés pour pondérer l'importance des différentes connexions, de sorte qu'une connexion peu importante contribue faiblement au déclenchement du neurone et vice versa. Le déclenchement du perceptron se décide par le calcul des entrées pondérées du neurone. La sortie binaire du neurone a_b , appelée activation, est donnée par l'équation suivante :

$$a_b = \begin{cases} 1 & \text{si } \mathbf{w}^\top \mathbf{x}_b > S \\ 0 & \text{si } \mathbf{w}^\top \mathbf{x}_b \leq S, \end{cases} \quad (3.1.1)$$

où S est le seuil du perceptron.

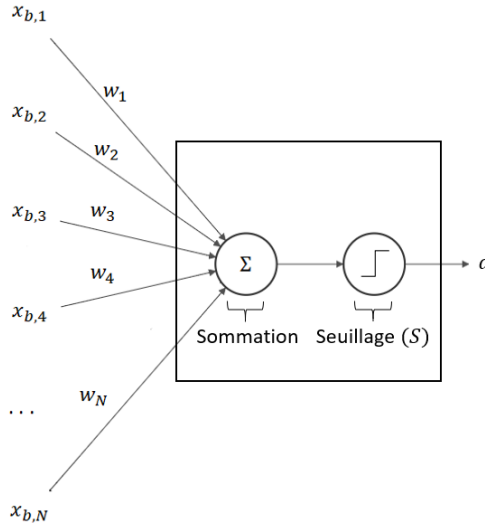


FIGURE 3.1. Le premier modèle du neurone artificiel. La série de N entrées binaires est envoyée à une unité de sommation et un seuil S est appliqué pour produire l'activation binaire a , suivant l'équation 3.1.1.

Le perceptron est d'abord modifié en lui attribuant un biais b au neurone. Le biais d'un neurone, défini par $b \equiv -S$, permet d'ajuster la façon avec laquelle un neurone peut s'activer. Les paramètres qui peuvent être ajustés pour modifier le comportement d'un neurone sont les poids de ses connexions et son biais.

De plus, plutôt que d'utiliser des entrées binaires, les neurones modernes utilisent des entrées réelles et le seuillage est remplacé par une fonction d'activation ξ . Une fonction d'activation est une fonction non linéaire et différentiable qui produit l'activation finale du neurone à partir de ses entrées, de ses poids et de son biais. L'utilisation des fonctions d'activation est essentielle pour permettre au réseau d'approximer des fonctions non linéaires. Une des justifications d'utiliser des activations réelles est qu'il est alors possible d'appliquer des algorithmes d'apprentissage, généralement basés sur une variante de la descente de gradient. L'apprentissage des paramètres et les fonctions d'activation sont discutés plus en détail dans la section 3.2.1.

Pour un neurone artificiel qui reçoit en entrée une donnée constituée de plusieurs valeurs réelles \mathbf{x} , l'activation a , qui est une valeur réelle, est donnée par l'équation suivante :

$$a = \xi \left(\mathbf{w}^\top \mathbf{x} + b \right). \quad (3.1.2)$$

Des réseaux de neurones peuvent être créés en faisant l'interconnexion de neurones artificiels. Une particularité intéressante des réseaux de neurones est qu'ils satisfont un théorème d'approximation universelle (UAT) [38–40]. En résumé, celui-ci stipule que toute fonction définie sur un sous-espace compact réel peut être approximée par un réseau de neurones avec une précision arbitraire. Toutefois, malgré l'existence de l'UAT, déterminer les paramètres du réseau ou en faire son implémentation peut être impossible, notamment à cause de l'absence d'algorithmes suffisamment puissants pour déterminer les paramètres. Néanmoins, couplés avec les algorithmes d'apprentissage existants pour déterminer automatiquement les paramètres, les réseaux de neurones sont des modèles particulièrement puissants. Ceux-ci sont abordés au cours de la prochaine section.

3.2. RÉSEAUX DE NEURONES

Un réseau de neurones est obtenu en connectant une série de neurones entre eux à l'aide de connexions pondérées. La figure 3.2 présente les différentes constituantes d'un réseau de neurones utilisé pour une classification à N_c classes distinctes. Les neurones sont organisés en une série de L couches, où les couches 1 et L correspondent respectivement aux couches d'entrée et de sortie. Chaque neurone de la couche d'entrée pourrait, par exemple, contenir l'intensité d'un voxel d'une région d'intérêt. La dernière couche ($l = L$) est constituée de N_c neurones et produit le vecteur de sortie \mathbf{a}^L , où la valeur a_i^L correspond à l'activation du neurone i , associé à la classe c_i . Les couches intermédiaires correspondent aux couches cachées du réseau. Pour une couche cachée l , chaque neurone est connecté à l'ensemble des neurones des couches $l - 1$ et $l + 1$.

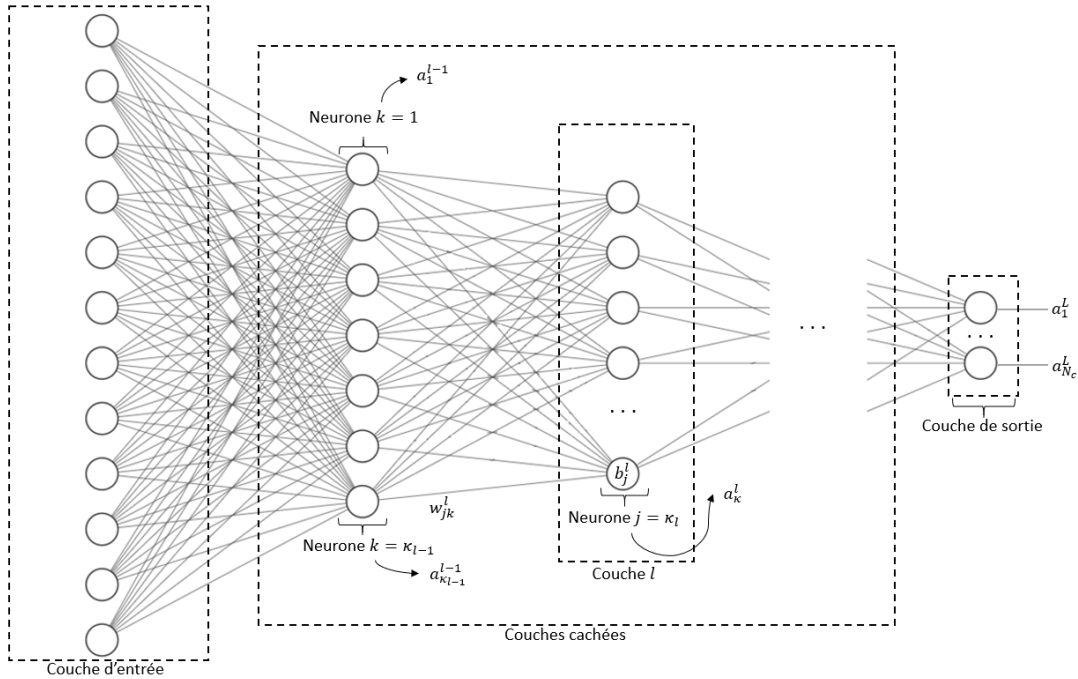


FIGURE 3.2. Un exemple de perceptron multicouche utilisé pour une classification à N_c classes distinctes. Les κ_l neurones qui composent la couche l sont connectés aux neurones des couches $l - 1$ et $l + 1$. La couche d'entrée contient un neurone par caractéristique extraite de la donnée d'entrée. Par exemple, 900 neurones d'entrée pourraient contenir les intensités des voxels d'une région d'intérêt de dimensions $30 \text{ px} \times 30 \text{ px}$. À partir de la couche d'entrée, les activations sont propagées vers l'avant pour produire en sortie les activations a_j^L utilisées pour faire la classification de la donnée d'entrée.

La majorité des types de réseaux de neurones, incluant les CNN discutés dans la section 3.4, sont à propagation avant. Dans ce type de réseau acyclique, l'information est propagée de la couche d'entrée vers la couche de sortie, contrairement aux réseaux de neurones récurrents. L'activation d'un neurone j d'une couche l est donnée par l'équation suivante :

$$a_j^l = \xi \left(\sum_{k=1}^{\kappa_{l-1}} w_{jk}^l a_k^{l-1} + b_j^l \right), \quad (3.2.1)$$

où $\sum_{k=1}^{\kappa_{l-1}} w_{jk}^l a_k^{l-1} + b_j^l$ est l'entrée pondérée du neurone j de la couche l , dénotée z_j^l .

L'équation 3.2.1 peut être utilisée avec un ensemble quelconque de poids et de biais à partir de la couche d'entrée afin de produire le vecteur d'activations a^L . Produire des a^L qui mènent à des prédictions valides implique d'ajuster les paramètres du réseau. La détermination des paramètres est généralement effectuée à l'aide d'un algorithme d'apprentissage supervisé.

3.2.1. Apprentissage

En apprentissage supervisé, il est nécessaire de posséder un ensemble de données dont les classes réelles sont connues. Celles-ci sont utilisées pour la détermination des paramètres lors d'une phase d'entraînement du réseau. Dans les problèmes de classification, la classe réelle peut être représentée par un vecteur de sortie \mathbf{y} . Ce vecteur est représenté en encodage *one-hot*, c'est-à-dire que la seule valeur non nulle se trouve à l'indice associé à la classe réelle et vaut 1. Par exemple, dans une classification à 3 classes distinctes $\{c_1, c_2, c_3\}$, une donnée ayant la classe réelle c_2 serait représentée par un vecteur $\mathbf{y} = (0 \ 1 \ 0)^T$. Les données disponibles peuvent être utilisées pour les activations d'entrée du réseau. La similarité entre les a^L obtenus par propagation avant et les \mathbf{y} qui y sont associés peut alors être évaluée par l'entremise d'une fonction coût J . L'évaluation de la fonction coût sert de point de départ pour l'utilisation d'un algorithme d'apprentissage.

3.2.1.1. Optimisation d'une fonction coût

La fonction coût J dépend du grand nombre de paramètres θ (poids et biais) du réseau. Celle-ci permet d'évaluer à quel point les prédictions du réseau sont près des classes réelles, aussi appelées réalités de terrain (*ground truth*). Les fonctions coût utilisées sont choisies de sorte à respecter deux critères afin d'être utilisées avec la rétropropagation (section 3.2.1.4). D'abord, celles-ci s'expriment comme une moyenne de fonctions coût J_x utilisées avec des données individuelles,

$$J(\theta) = \frac{1}{m} \sum_{j=1}^m J_{x,j}(\theta), \quad (3.2.2)$$

où m est la taille de mini-lot, c'est-à-dire la taille du sous-ensemble des données d'entraînement considéré par J . Aussi, J dépend uniquement des activations des neurones de la couche de sortie. Les principales fonctions coûts utilisées avec les réseaux de neurones sont décrites dans la section 3.2.1.6.

La minimisation de J se fait généralement par une variante de l'algorithme de descente de gradient. En résumé, la descente de gradient est une méthode itérative qui consiste à se déplacer d'un pas proportionnel au gradient, dans la direction opposée à celui-ci, de façon itérative. Pour les réseaux de neurones, cela implique de calculer $\nabla_{\theta} J$ et de modifier itérativement les poids et les biais. Pour un paramètre θ , l'équation considérée lors d'une descente de gradient est donnée par :

$$\theta_{i+1} = \theta_i - \frac{\eta}{m} \sum_{j=1}^m \frac{\partial J_{x,j}(\theta)}{\partial \theta_i}, \quad (3.2.3)$$

où η est le pas de l'itération, appelé taux d'apprentissage. Le taux d'apprentissage est généralement choisi comme étant inférieur à 1, de sorte à avoir de petites itérations lors de la descente de gradient, mais assez élevé pour effectuer l'apprentissage dans un temps réaliste. Pour déterminer une valeur convenable de η , il est possible, par exemple, de débiter avec une valeur relativement élevée et d'en faire la décroissance progressive lors de l'entraînement.

La descente de gradient classique implique de calculer le gradient de J_x pour chacune des données d'entraînement. Cependant, une quantité importante de données est généralement utilisée et le calcul des gradients peut être très long.

Une première variante, soit la descente de gradient stochastique (SGD), introduit l'utilisation de mini-lots de données pour rendre réaliste le temps d'entraînement. Les paramètres sont alors optimisés à partir d'une approximation de ∇J , en calculant les gradients individuels pour un petit sous-ensemble des données d'entraînement sélectionnées aléatoirement. Tant que toutes les données d'entraînement n'ont pas été utilisées dans un mini-lot, de nouveaux sous-ensembles, mutuellement exclusifs, sont générés pour entraîner le réseau. Lorsque toutes les données d'entraînement ont été envoyées une fois au réseau, une époque d'entraînement est terminée. L'entraînement du réseau peut être effectué pendant plusieurs époques. Actuellement, de nombreux algorithmes d'optimisation se basent sur la SGD. Les deux algorithmes d'optimisation parmi les plus utilisés en apprentissage profond sont discutés dans les prochaines sous-sections.

3.2.1.2. Descente de gradient stochastique avec momentum

Malgré que la SGD puisse être utilisée dans de nombreux cas, l'apprentissage peut être très lent [41]. Une variante de l'algorithme, la descente de gradient stochastique avec momentum (SGDM), permet de converger vers un minimum beaucoup plus rapidement. Le principe de la SGDM est analogue à une particule classique qui se déplacerait dans l'espace des paramètres. Un vecteur vitesse \mathbf{v} est introduit et les modifications apportées par le gradient sont appliquées sur \mathbf{v} . Un terme tient compte des gradients des itérations précédentes et celui-ci est pondéré par un coefficient de momentum f . Le coefficient de momentum est un scalaire prenant une valeur entre 0 et 1. Un coefficient de 0 correspond au cas où il n'y a aucune contribution de l'itération précédente, tandis qu'il y a une contribution importante de l'itération précédente pour $f = 1$. La SGDM utilise les équations suivantes [41] :

$$\mathbf{v}_{i+1} = f\mathbf{v}_i - \frac{\eta}{m}\nabla_{\boldsymbol{\theta}_i}J(\boldsymbol{\theta}_i) \quad (3.2.4)$$

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \mathbf{v}_{i+1} \quad (3.2.5)$$

Toutefois, l'algorithme peut rencontrer des problèmes dans des situations pathologiques telles qu'à la rencontre d'un point de selle (*saddle point*), où la convergence peut être très lente. Ceci étant dit, l'algorithme offre en général de bonnes performances et est robuste dans la majorité des espaces des paramètres.

3.2.1.3. Estimation adaptative de moments

D'autres optimiseurs ont également été développés, notamment les méthodes adaptatives. Celles-ci incluent l'estimation adaptative de moments (Adam), reconnue pour sa robustesse dans l'optimisation d'objectifs stochastiques dans les espaces de paramètres à haute dimension [42]. L'algorithme de l'optimiseur Adam est présenté à l'algorithme 1. Contrairement au SGDM qui utilise un seul taux d'apprentissage, l'optimiseur Adam calcule différents taux d'apprentissage qui s'adaptent pour les différents paramètres. Ceux-ci sont basés sur des estimations du premier et du deuxième moment du gradient, u_1 et u_2 respectivement, corrigées pour le biais engendré par l'initialisation à 0 des vecteurs. L'optimiseur introduit des taux de décroissance, β_1 et β_2 , présents dans l'estimation des moments, ainsi qu'une constante $\epsilon \ll 1$ présente dans le calcul de $\Delta\theta$ pour assurer une stabilité numérique.

Algorithme 1 Optimiseur ADAM

Require: Hyper-paramètres : $\beta_1, \beta_2, \epsilon, \eta, m$

Require: Paramètres initiaux θ

```

1:  $u_{1,0} \leftarrow 0$                                  $\triangleright$  Initialisation du vecteur des premiers moments
2:  $u_{2,0} \leftarrow 0$                                  $\triangleright$  Initialisation du vecteur des deuxièmes moments
3:  $t \leftarrow 0$                                      $\triangleright$  Initialisation du compteur d'itérations
4: while Convergence de  $\theta_t$  non atteinte do
5:   Mini-lot : échantillonner  $m$  données de l'ensemble d'entraînement
6:    $t \leftarrow t + 1$ 
7:    $\mathbf{G}_t \leftarrow \nabla_{\theta} J(\theta_{t-1})$                  $\triangleright$  Calcul du gradient de la fonction coût
8:    $u_{1,t} \leftarrow \beta_1 u_{1,t-1} + (1 - \beta_1) \mathbf{G}_t$      $\triangleright$  Estimation du premier moment
9:    $u_{2,t} \leftarrow \beta_2 u_{2,t-1} + (1 - \beta_2) \mathbf{G}_t \odot \mathbf{G}_t$   $\triangleright$  Estimation du deuxième moment
10:   $\hat{u}_{1,t} = \frac{u_{1,t}}{1 - \beta_1^t}$                  $\triangleright$  Correction du biais du premier moment
11:   $\hat{u}_{2,t} = \frac{u_{2,t}}{1 - \beta_2^t}$                  $\triangleright$  Correction du biais du deuxième moment
12:   $\Delta\theta_t = -\eta \frac{\hat{u}_{1,t}}{\sqrt{\hat{u}_{2,t} + \epsilon}}$      $\triangleright$  Calcul des corrections à appliquer aux paramètres
13:   $\theta_t \leftarrow \theta_{t-1} + \Delta\theta_t$                  $\triangleright$  Modification des paramètres
14: end while
```

3.2.1.4. Rétropropagation

Dans les sections précédentes, l'entraînement des réseaux de neurones a été abordé. Malgré les nombreuses variantes d'algorithmes d'optimisation, celles-ci ont en commun le calcul de $\nabla_{\theta} J$. Pour apporter les modifications aux poids et

aux biais du réseau, la méthode généralement utilisée dans le calcul des gradients est la rétropropagation, qui est abordée dans cette section.

La rétropropagation peut être effectuée en introduisant une quantité intermédiaire δ_j^l , qui correspond à l'erreur d'un neurone [43]. L'erreur d'un neurone peut être interprétée comme une évaluation de sa contribution à la fonction coût. Plus l'erreur est élevée, plus l'entraînement devra modifier de façon importante les paramètres du neurone pour qu'il contribue de façon favorable aux prédictions. Pour un neurone j de la couche l , l'erreur peut être définie en fonction de l'entrée pondérée z_j^l avec l'équation suivante [43] :

$$\delta_j^l \equiv \frac{\partial J}{\partial z_j^l}. \quad (3.2.6)$$

La rétropropagation consiste d'abord à calculer l'erreur des neurones de la couche de sortie, δ^L . Ensuite, les δ_j^l et toutes des composantes du gradient peuvent être calculées avec des relations de récurrence pour l'ensemble du réseau. Des équations permettant de calculer les composantes du gradient sont présentées aux équations 3.2.7 à 3.2.10 [43].

$$\delta_j^L = \frac{\partial J}{\partial a_j^L} \frac{\partial \xi(z_j^L)}{\partial z_j^L} \quad (3.2.7)$$

$$\delta_j^l = \sum_k w_{kj}^{l+1} \delta_k^{l+1} \frac{\partial \xi(z_j^l)}{\partial z_j^l} \quad (3.2.8)$$

$$\frac{\partial J}{\partial b_j^l} = \delta_j^l \quad (3.2.9)$$

$$\frac{\partial J}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l \quad (3.2.10)$$

La fonction coût dépend uniquement des vecteurs a^L et \mathbf{y} , donc l'équation 3.2.7 peut être utilisée comme point de départ à partir des prédictions faites sur les données d'entraînement. Ensuite, l'équation 3.2.8 permet de faire la rétropropagation de l'erreur dans l'ensemble du réseau et les équations 3.2.9 et 3.2.10 permettent le calcul des composantes du gradient. En résumé, la combinaison d'un algorithme d'optimisation tel que l'optimiseur Adam avec la rétropropagation permet d'entraîner un réseau de neurones à partir d'un ensemble de données d'entraînement.

3.2.1.5. Fonctions d'activation

Comme mentionné précédemment, l'utilisation d'une fonction d'activation non linéaire est essentielle pour permettre au réseau d'approximer des fonctions non linéaires. Par exemple, au début de ce chapitre, à l'équation 3.1.1, l'activation est donnée par l'utilisation d'une fonction de Heaviside, qui permet de donner en sortie une activation de 0 ou de 1. L'utilisation des fonctions d'activation est également une condition à l'application du théorème d'approximation universelle. Dans cette section, les principales fonctions d'activation sont introduites. Celles-ci sont également présentées à la figure 3.3. Une des premières fonctions d'activation utilisée est la fonction sigmoïde, donnée par l'équation suivante :

$$\xi_{\text{sig}}(z_j^l) = \frac{1}{1 + e^{-z_j^l}}. \quad (3.2.11)$$

Comme avec toutes les fonctions présentées dans cette section, l'utilisation de la sigmoïde comme fonction d'activation permet d'obtenir de bonnes performances. Cependant, des problèmes de disparition de gradient et de ralentissement d'apprentissage peuvent survenir. Ces problèmes, nuisibles à l'entraînement, sont principalement dus à la forme de la courbe pour les valeurs extrêmes de z_j^l ainsi qu'à sa dérivée. Par exemple, un neurone sigmoïde pourrait avoir un z_j^l extrême si un biais élevé lui est attribué lors de l'initialisation. Comme l'erreur du neurone restera très petite, les modifications apportées aux paramètres du neurone le seront également. Une situation de neurones morts (*dying neurons*) peut alors survenir, c'est-à-dire que pendant l'entraînement, les paramètres de certains neurones ne peuvent pas être modifiés de manière appréciable.

Dans le même ordre d'idées, la tangente hyperbolique est aussi utilisée comme fonction d'activation, donnée par l'équation suivante :

$$\xi_{\text{tanh}}(z_j^l) = \frac{e^{z_j^l} - e^{-z_j^l}}{e^{z_j^l} + e^{-z_j^l}}. \quad (3.2.12)$$

Celle-ci peut prendre des valeurs se situant dans l'intervalle $[-1, 1]$ et l'activation sera du même signe que l'entrée pondérée. Toutefois, étant donné la relation entre la sigmoïde et la tangente hyperbolique ($\xi_{\tanh}(z_j^l) = 2\xi_{\text{sig}}(2z_j^l) - 1$), celle-ci présente des inconvénients similaires à la sigmoïde.

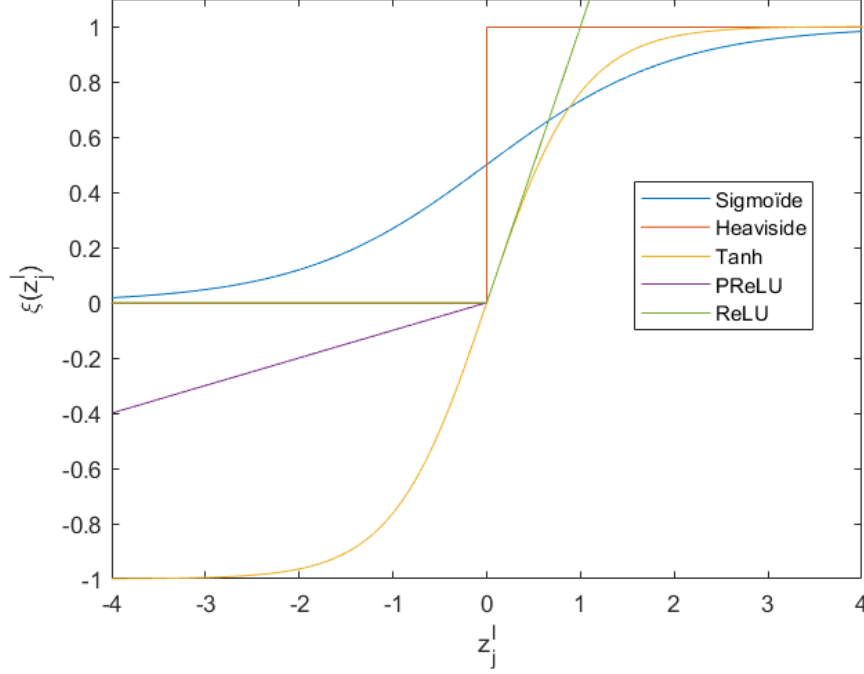


FIGURE 3.3. Graphique des principales fonctions d'activation (sigmoïde, tangente hyperbolique, unités linéaires rectifiées non paramétrique (ReLU) et paramétrique (PReLU), fonction échelon).

Un groupe de fonctions d'activation largement utilisé est la famille des unités linéaires rectifiées (ReLU) [44]. Ces fonctions peuvent améliorer considérablement les performances des modèles et elles sont généralement plus performantes que la sigmoïde et la tangente hyperbolique dans le contexte des réseaux neuronaux [45, 46]. Un avantage majeur est que les unités linéaires rectifiées accélèrent grandement l'entraînement des réseaux, ce qui permet de travailler avec des réseaux plus complexes [47]. La fonction ReLU est donnée par l'équation suivante :

$$\text{ReLU}(z_j^l) = \max(0, z_j^l). \quad (3.2.13)$$

Malgré les avantages de la fonction ReLU, la possibilité d'avoir des neurones morts est toujours présente. Des variantes ont été proposées pour corriger ce problème tout en gardant les avantages de la fonction ReLU. Un exemple est la ReLU avec fuite (LReLU), donnée par l'équation suivante [48] :

$$\text{LReLU} \left(z_j^l \right) = \max \left(0, z_j^l \right) + \alpha \min \left(0, z_j^l \right), \quad (3.2.14)$$

où α est un paramètre à déterminer pour la pente du terme de fuite. Une autre variante est la ReLU paramétrique (PReLU), également donnée par l'équation 3.2.14, à la différence où α est déterminé pendant l'entraînement [49].

Une autre fonction d'activation particulièrement importante dans les problèmes de classification et de segmentation est la fonction softmax. La fonction softmax est utilisée par les neurones de la couche de sortie. En prenant comme argument les entrées pondérées z^L , les activations résultantes de la fonction softmax s'interprètent comme des probabilités. Pour le neurone i de la couche de sortie, l'activation softmax So est donnée par l'équation 3.2.15 et correspond à la probabilité du neurone i d'être de la classe associée à l'indice n_c , connaissant z^L .

$$So \left(z^L \right) = \frac{e^{z_i^L}}{\sum_{n_c=1}^{N_c} e^{z_{n_c}^L}} \quad (3.2.15)$$

Ensuite, la classe associée à l'activation softmax la plus élevée peut être attribuée à la donnée d'entrée. Aussi, tel que discuté précédemment, ce vecteur est envoyé à la fonction coût lors de la phase d'entraînement.

3.2.1.6. Fonctions de coût

L'entraînement du réseau de neurones repose sur la minimisation d'une fonction coût J . En minimisant J , l'objectif est également d'améliorer des métriques de performances, telles que la précision, la sensibilité et le coefficient Dice (voir le chapitre 4). Une première fonction coût pouvant être utilisée est l'erreur quadratique moyenne (MSE), donnée par l'équation suivante :

$$J = \frac{1}{2m} \sum_{i=1}^m \sum_{j=1}^{N_c} \left(a_{j,i}^L - y_{j,i} \right)^2. \quad (3.2.16)$$

Avec la MSE, les écarts entre les activations de chaque neurone de sortie et la sortie attendue sont sommés pour toutes les données d'un mini-lot. Celle-ci peut être utilisée avec la rétropropagation. Cependant, elle présente l'inconvénient

d'introduire une dépendance en $\frac{\partial \xi(z_j^L)}{\partial z_j^L}$ dans les composantes du gradient des neurones de sortie qui servent de point de départ à la rétropropagation. La dépendance sur la fonction d'activation peut causer des ralentissements d'apprentissage, par exemple, lorsque la fonction sigmoïde est utilisée. En général, il est souhaitable d'avoir une fonction coût qui, lorsque combinée avec la fonction d'activation, permet de calculer des δ_j^L indépendants de $\xi'(z_j^L)$.

Une fonction coût particulièrement intéressante est l'entropie croisée, donnée par l'équation suivante :

$$J = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{N_c} y_{j,i} \ln a_{j,i}^L. \quad (3.2.17)$$

Tel que mentionné à la section 3.2.1.5, les propriétés de la fonction softmax en font une fonction d'activation très intéressante pour les neurones de sortie. En utilisant conjointement la fonction softmax et l'entropie croisée, l'erreur des neurones de sortie dépend de l'écart entre l'activation et la sortie attendue et non de la forme de la fonction d'activation [43].

3.3. SURAPPRENTISSAGE

Pendant l'entraînement d'un réseau, ses paramètres sont déterminés itérativement à l'aide de données et d'algorithmes d'apprentissage. Idéalement, durant l'optimisation des paramètres, un réseau entraîné aura appris à extraire des caractéristiques qui se généralisent bien à d'autres données du domaine d'application. Par exemple, un réseau entraîné sur des images contenant du verre dépoli devrait avoir appris à associer une augmentation de la densité sur un tissu peu texturé à du verre dépoli.

Le nombre de paramètres dans un réseau est généralement très élevé, pouvant atteindre des millions de poids et de biais. Ainsi, l'entraînement implique d'avoir à sa disposition une énorme quantité de données, souvent inaccessible. Ce problème est amplifié en imagerie médicale, où les données sont limitées. L'apprentissage supervisé est alors effectué avec les données disponibles pendant plusieurs époques (itérations sur tout l'ensemble de données). Toutefois, itérer sur peu de données pendant un grand nombre d'époques peut causer un surapprentissage (*overfitting*). Cela signifie qu'en plus des caractéristiques

souhaitables, un réseau en surapprentissage aura appris à chercher des particularités et du bruit propres aux données d'entraînement qui ne se généralisent pas à de nouvelles données.

Par conséquent, sa capacité de généralisation sera plus faible. Cela se traduit par des performances considérablement meilleures sur les données d'entraînement que sur les données de test, exclues de la phase d'entraînement. Afin d'éviter le surapprentissage, il est possible de réduire le nombre de paramètres du réseau, par exemple, en réduisant le nombre de couches cachées ou en diminuant la quantité de neurones des couches existantes. Toutefois, diminuer la complexité du réseau peut avoir un impact négatif sur les performances atteintes. Déterminer des dimensions convenables pour un réseau utilisé dans une application donnée demande, entre autres, d'évaluer la complexité du problème étudié et les données disponibles. Par exemple, le nombre de classes considérées peut avoir un impact sur la quantité de données et la taille du réseau nécessaires. Pour donner quelques ordres de grandeur, les CNN les plus complexes actuellement utilisés en imagerie peuvent avoir plus de 100 couches, dans un contexte de classification d'images de 1000 classes distinctes. De tels réseaux sont généralement entraînés avec plus d'un million d'images. Dans le cadre d'applications plus restreintes, où relativement peu de classes sont considérées, d'excellentes performances peuvent être obtenues avec des réseaux de quelques dizaines de couches et un ensemble de données de quelques centaines d'images.

Avant de considérer la possibilité de réduire la complexité d'un réseau, la mise en place de techniques de régularisation durant l'entraînement est souvent suffisante pour éliminer ou grandement diminuer le surapprentissage. De façon générale, une technique de régularisation vise à apporter des modifications à l'algorithme d'apprentissage dans le but de réduire l'erreur de généralisation du modèle, sans nécessairement réduire son erreur durant l'entraînement [41]. Pour ce faire, la régularisation vise à ajouter des perturbations qui sont uniquement présentes durant l'entraînement.

Des exemples de techniques de régularisation incluent l'arrêt prématuré (*early stopping*), l'augmentation de données, la dégradation des poids (*weight decay*) et l'abandon (*dropout*). Ces techniques sont brièvement abordées dans cette section. Avec la combinaison de ces techniques de régularisation, l'entraînement de réseaux beaucoup plus complexes devient possible, même avec une quantité limitée de données.

3.3.1. Arrêt prématuré

L'arrêt prématuré consiste à effectuer l'entraînement du réseau jusqu'au point où les performances sur des données inconnues commencent à diminuer [43]. Comme les données de test sont uniquement utilisées à la fin de l'entraînement pour évaluer les performances finales, un troisième ensemble de données, dit de validation, est intégré. Ceci implique donc de réduire la quantité de données utilisées durant l'entraînement et les tests. Tout comme avec les données de test, l'ensemble de validation est uniquement utilisé pour évaluer les performances. Après un certain nombre d'époques d'entraînement, une évaluation des données de validation est faite et les performances sont calculées. Lorsque le réseau devient moins performant sur les données de validation, l'optimisation des paramètres est alors arrêtée.

3.3.2. Augmentation de données

Une autre technique consiste à faire une augmentation artificielle de la quantité de données disponibles dans l'ensemble d'entraînement. Pour ce faire, des transformations peuvent être appliquées aux données avant d'être envoyées au réseau [50]. En imagerie, des transformations géométriques simples (translation, rotation, réflexion) peuvent être appliquées aléatoirement. D'autres exemples consistent à modifier la luminosité et à ajouter du bruit. Dans certains cas, il peut également être pertinent d'ajouter des transformations propres à l'application considérée. Un exemple intéressant est l'augmentation de données utilisée pour le problème classique de classification de chiffres manuscrits de la base de données modifiée du National Institute of Standards and Technology (NIST) [51]. Pour cette application, une étude a augmenté ses données d'entraînement en intégrant des distorsions élastiques qui simulent les oscillations aléatoires dans les muscles de la main et les performances ont augmenté considérablement [52].

Selon le type de données considéré, une attention particulière doit tout de même être portée aux types de transformations utilisées. Dans certains cas, l'augmentation de données peut produire des données d'entraînement qui ont des caractéristiques non représentatives des images réelles. Un exemple de cette situation serait d'ajuster fortement la luminosité d'une image CT qui contient du verre dépoli. Le schéma de la figure 3.4 présente une façon d'intégrer l'augmentation de données à l'entraînement. De nouvelles transformations peuvent être appliquées à chaque époque de sorte à diversifier davantage les données d'entraînement.

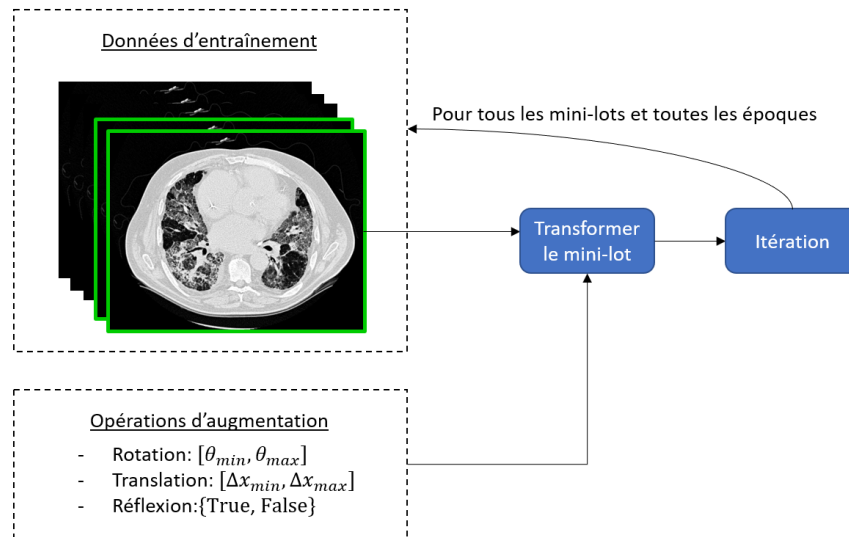


FIGURE 3.4. L'augmentation de données peut être intégrée à l'entraînement d'un réseau pour augmenter artificiellement le volume de données d'entraînement. Une façon de procéder à l'augmentation de données est de sélectionner aléatoirement des transformations à appliquer parmi des opérations et intervalles prédéfinis. Les données d'entraînement du mini-lot considéré (encadrées en vert) peuvent ensuite être transformées et l'entraînement est effectué avec ce mini-lot. Cette étape d'entraînement avec augmentation est ensuite répétée avec tous les mini-lots, et ce, pour toutes les époques d'entraînement.

3.3.3. Dégradation des poids

Une autre technique de régularisation qui contribue à éviter le surapprentissage est la dégradation des poids. Celle-ci consiste à apporter une modification à la fonction coût en y intégrant un terme de régularisation qui dépend uniquement du tenseur de poids [53]. Le terme ajouté correspond à une pénalité sur la norme des poids $\Omega(\mathbf{w})$. Intégrer ce terme à la fonction coût utilisée pendant l'entraînement implique, qu'en plus de chercher à minimiser une fonction telle que l'entropie croisée, la valeur des poids sera aussi diminuée. La forme de la fonction coût régularisée \tilde{J} est donnée par l'équation suivante [41] :

$$\tilde{J} = J(\boldsymbol{\theta}) + \lambda \Omega(\mathbf{w}), \quad (3.3.1)$$

où λ est le facteur de régularisation. Généralement, pour les réseaux de neurones, le terme de pénalité utilisé est basé sur la régularisation L^2 et est donné par $\Omega(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}$. En résumé, cette méthode de dégradation des poids a pour effet de fortement pénaliser les configurations où quelques neurones contribueraient de façon importante aux activations finales à cause de poids importants. Au contraire, des solutions où tous les neurones contribuent dans des proportions comparables seront favorisées [54]. Une alternative parfois utilisée est l'utilisation d'une pénalité basée sur la norme L_1 , donnée par $\Omega(\mathbf{w}) = \sum_i |w_i|$. Comparativement à la régularisation L^2 , cette alternative a tendance à produire des solutions creuses (*sparse*), c'est-à-dire qu'une partie considérable du tenseur de poids final sera approximativement nulle. Pour cette raison, la régularisation L_1 est davantage utilisée dans les problèmes de sélection de caractéristiques (*feature selection*) [41].

3.3.4. Abandon

L'agrégation de bootstrap (*bagging*, *bootstrap aggregating*) consiste à combiner plusieurs modèles dans le but de réduire l'erreur de généralisation [41]. Une technique simple serait de faire voter plusieurs modèles entraînés sur chaque donnée de test. Toutefois, cela implique d'entraîner plusieurs réseaux et les ressources nécessaires peuvent rapidement devenir excessives. L'abandon (*dropout*) est une approximation peu coûteuse au *bagging* qui peut être utilisée comme technique de régularisation dans les réseaux de neurones afin d'avoir un entraînement plus robuste [55]. Durant l'entraînement, l'abandon consiste à

désactiver aléatoirement certains neurones de sorte que chaque donnée d'entraînement soit analysée par un sous-ensemble de l'architecture initiale. À chaque inférence (passage d'une donnée dans le réseau), un sous-réseau différent est utilisé. Un exemple d'utilisation de l'abandon est présenté à la figure 3.5.

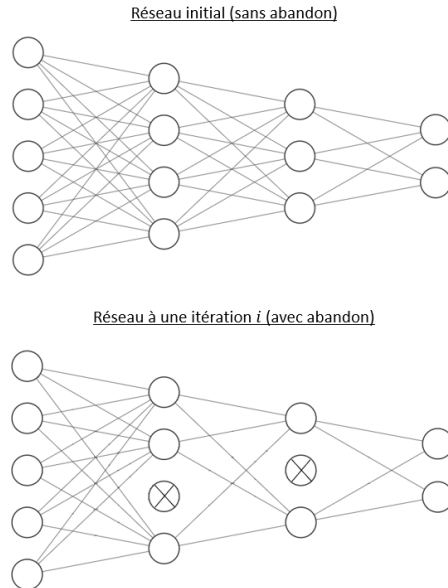


FIGURE 3.5. Le réseau connecté du haut montre une architecture qui doit passer par une phase d'entraînement. En intégrant l'abandon de neurones à l'entraînement, certains neurones sont désactivés aléatoirement à chaque passage dans le réseau, et un sous-réseau effectif différent est utilisé à chaque itération (par exemple, le réseau du bas).

3.4. RÉSEAUX DE NEURONES À CONVOLUTION

Dans les sections précédentes, un survol des réseaux de neurones et de l'apprentissage supervisé a été fait. De plus, le problème de surapprentissage et différentes méthodes permettant de le réduire ont été abordés. Tous ces concepts sont pertinents pour les réseaux profonds (à plusieurs couches cachées). Cependant, les réseaux de neurones connectés (perceptron multicouche, MLP) ne sont généralement pas convenables pour développer des réseaux complexes, où le nombre de couches et de paramètres sont énormes. En effet, certains des fondements des MLP mènent à des problèmes importants qui limitent les performances, notamment la disparition du gradient [43, 56, 57]. En résumé, à chaque couche traversée, le gradient rétropropagé peut devenir de plus en plus

petit au point où l'apprentissage des premières couches ne se fait pas correctement. Aussi, une limitation du MLP est l'explosion du nombre de paramètres causée par les connexions totales. Cela impose une limite sur le nombre de neurones d'entrée, la profondeur du réseau et sa complexité. En imagerie, la taille des images brutes pouvant être envoyées au réseau est fortement limitée. Aussi, ces réseaux n'utilisent pas l'information spatiale présente dans les images, alors que cette information est extrêmement bénéfique pour leur compréhension.

Le réseau de neurones à convolution (CNN) est une variante du MLP qui permet d'apporter des solutions aux limitations discutées. Les CNN sont inspirés par le système visuel biologique et utilisent l'information spatiale présente dans les données. L'intégration de l'information spatiale dans le traitement effectué permet d'obtenir des performances particulièrement intéressantes en analyse d'images. Aussi, les connexions entre les neurones des différentes couches ne sont pas totales, ce qui permet l'entraînement de réseaux beaucoup plus complexes. Au cours des prochaines sections, les CNN ainsi que leur utilisation en imagerie et en segmentation sémantique sont abordés.

3.4.1. Paramètres et activations

Tout comme les perceptrons multicouches, les CNN ont des poids et des biais déterminés durant une phase d'entraînement. Cependant, tel que discuté dans les prochaines sous-sections, la façon dont les CNN sont structurés réduit grandement le nombre de paramètres à déterminer, comparativement aux perceptrons multicouches où les connexions sont totales. Aussi, les modifications au calcul des activations dans les CNN sont abordées.

3.4.1.1. Champ récepteur local et partage de paramètres

D'abord, les neurones sont organisés en volumes plutôt que linéairement. La figure 3.6 schématise l'organisation des neurones dans deux couches adjacentes. Pour une couche l , ils sont disposés en H_l lignes et en W_l colonnes, de façon similaire aux valeurs d'intensité d'une image, ainsi qu'en $N_{a,l}$ cartes d'activations (*feature maps*) dénotées $a_{i,l}$. Pour la couche d'entrée, les différentes cartes correspondent généralement aux couleurs de base servant à représenter l'image. Pour une image en niveaux de gris, une seule carte ($N_{a,1} = 1$) peut être

utilisée. Pour les images couleur, on peut utiliser trois cartes pour les couleurs rouge, vert et bleu. Pour les couches cachées, l'information est également représentée sous forme de cartes d'activations. Chacune de ces cartes est utilisée pour représenter une caractéristique (*feature*) particulière de l'image. En imagerie, une caractéristique correspond à de l'information extraite de l'image utilisée pour la compréhension de celle-ci. Des exemples de caractéristiques incluent la présence de certains contours, textures ou objets.

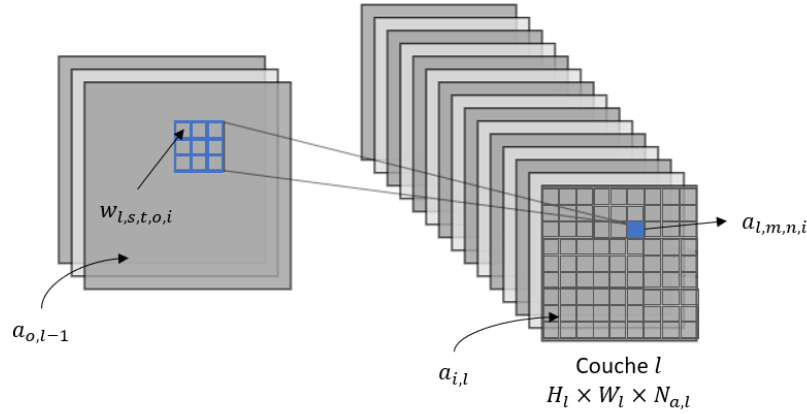


FIGURE 3.6. Deux couches illustrant l'organisation des neurones et des connexions dans un CNN. Chaque couche l contient des neurones organisés en H_l lignes, en W_l colonnes et en une série de $N_{a,l}$ cartes d'activations. Le neurone situé à la couche l , ligne m , colonne n et carte d'activation i ($a_{l,m,n,i}$) est partiellement connecté à une fenêtre de neurones de la couche précédente (quadrillé bleu). À l'intérieur de cette fenêtre, tous les poids associés à la carte i de la couche l interviennent dans le calcul de l'activation $a_{l,m,n,i}$. Dans la matrice de poids, un poids peut être localisé par une ligne s , une colonne t et un indice de carte o ($w_{l,s,t,o,i}$).

Une autre particularité des CNN est l'utilisation de connexions partielles. Un exemple de neurone partiellement connecté est également représenté à la figure 3.6. Alors que les neurones des MLP sont connectés avec tous les neurones des couches adjacentes, ceux des CNN sont uniquement connectés à un sous-ensemble des neurones adjacents. L'activation d'un neurone est directement déterminée par les activations des neurones auxquels il est connecté, son biais et les poids des connexions. Les connexions partielles impliquent que les neurones sont uniquement affectés par une région de l'image d'entrée et non la totalité de

celle-ci. La région de l'image qui est vue par un neurone correspond à son champ récepteur. Plus un neurone fait partie d'une couche profonde, plus son champ récepteur sera élevé. Sommairement, l'information extraite avec un petit champ récepteur peut être davantage interprétée comme de l'information spatiale, alors qu'un grand champ récepteur contient essentiellement de l'information sémantique. Augmenter progressivement le champ récepteur dans le réseau permet de considérer plusieurs niveaux d'abstractions et échelles de façon hiérarchique. Différentes opérations sont utilisées dans le but de contrôler l'augmentation du champ récepteur (voir la section 3.4.2).

Aussi, les paramètres sont partagés pour tous les neurones d'une carte $a_{l,i}$. En effet, alors que tous les neurones d'une couche de MLP ont des poids différents, un seul tenseur de poids $w_{l,i}$ est associé à tous les neurones d'une carte d'activations et, de façon similaire, un seul biais $b_{l,i}$ est partagé. L'application du $w_{l,i}$ d'une carte d'activations, analogue à un filtre, sur les différentes cartes de la couche précédente permet d'extraire une caractéristique particulière. Comme avec les MLP, les activations peuvent ensuite être générées en ajoutant le biais et en appliquant une fonction d'activation.

3.4.1.2. Cartes d'activations

Dans les CNN, les activations d'une carte sont basées sur une opération de corrélation croisée discrète. Pour une carte i de la couche l , l'opération est effectuée entre la totalité des activations représentées à la couche précédente et le filtre $w_{l,i}$. Pour un neurone (m,n) d'une carte (l,i) , une hauteur de filtre de K_1 et une largeur de filtre de K_2 , l'activation peut être calculée avec l'équation suivante :

$$a_{l,m,n,i} = \xi \left(b_{l,i} + \sum_{o=1}^{N_{a,l-1}} \sum_{t=1}^{K_2} \sum_{s=1}^{K_1} w_{l,s,t,o,i} a_{l-1,m+s,n+t,o} \right) \quad (3.4.1)$$

Donc, pour une carte entière, les activations peuvent être déterminées avec l'équation suivante :

$$a_{l,i} = \xi (b_{l,i} + w_{l,i} * a_{l-1}), \quad (3.4.2)$$

où $*$ dénote l'opération de corrélation croisée. La corrélation croisée est privilégiée puisqu'elle fournit la même information qu'une convolution, mais aucune rotation du noyau n'est effectuée, ce qui accélère le calcul des activations. Le terme convolution est tout de même utilisé pour garder une

cohérence avec la terminologie utilisée dans le domaine. Comme il est possible de le voir à l'équation 3.4.1, chacune des cartes d'activations possède un filtre à 3 dimensions. La troisième dimension des filtres, dénotée par le compteur o , représente les différentes cartes de la couche précédente. Les poids d'une couche peuvent être représentés par un tenseur de dimension $K_1 \times K_2 \times N_{a,l-1} \times N_{a,l}$, de sorte que l'équation 3.4.2 génère des cartes bidimensionnelles. Des exemples de cartes d'activations sont présentés à la figure 3.7.

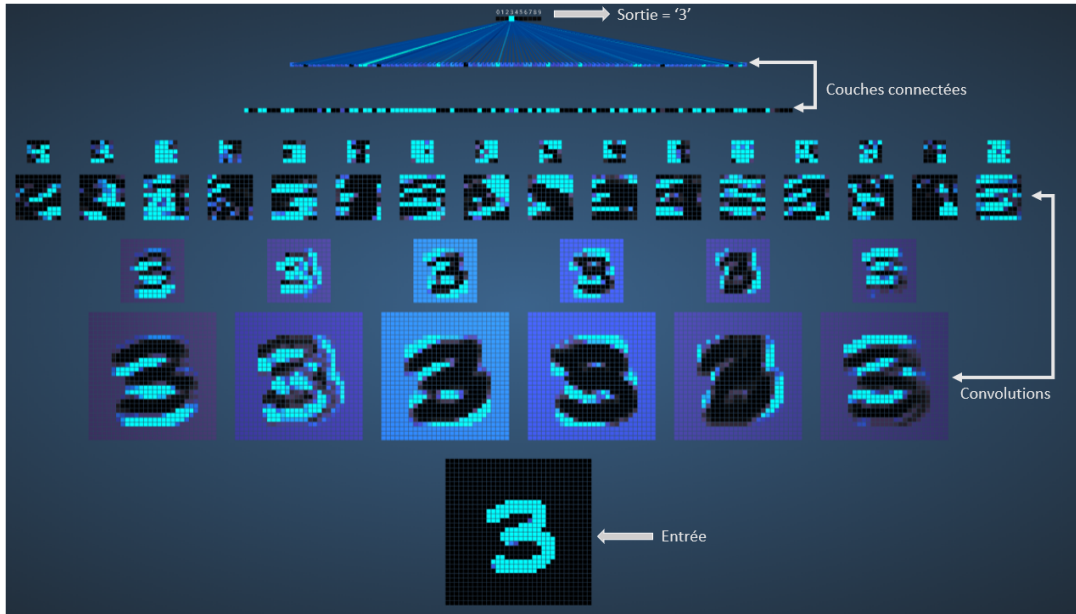


FIGURE 3.7. Un chiffre manuscrit est donné en entrée à un CNN contenant 2 couches de convolution, 2 couches de sous-échantillonnage et 2 couches connectées (voir la section 3.4.2). Les neurones de la première couche de convolution ont un petit champ récepteur et les cartes représentent de l'information de bas niveau (contours). Les neurones de la deuxième couche se basent sur les cartes de la première couche pour générer des cartes plus abstraites. Les cartes sont envoyées aux 2 couches connectées, qui éliminent l'information spatiale, produisent une activation qui permet d'attribuer une classe au chiffre manuscrit d'entrée. Les activations représentées en cyan correspondent à des activations plus élevées. Cartes d'activations générées à partir de <http://scs.ryerson.ca/~aharley/vis/>.

3.4.2. Types d'opérations et de couches

Les architectures de CNN sont constituées de plusieurs types de couches, telles que les couches de convolution et les couches d'augmentation de champ récepteur. Des couches connectées peuvent également être présentes à la fin du réseau. Dans cette section, les principales couches et opérations pouvant être retrouvées dans un CNN sont introduites.

3.4.2.1. *Convolution*

Les couches de convolution sont les composantes de base des CNN. Une telle couche est organisée en une série de cartes dont les activations sont données par l'équation 3.4.2. Avec l'entraînement du réseau, les différents filtres des couches de convolution, ainsi que les biais, peuvent être optimisés. Pour chacune des couches de convolution, plusieurs paramètres tels que les dimensions du noyau de convolution et le nombre de cartes d'activations à produire en sortie sont à déterminer. Généralement, les petits noyaux sont préférables pour la détection des détails plus fins qui exigent une bonne résolution. La taille du champ récepteur peut alors être augmentée progressivement dans la suite du réseau. Aussi, lorsque les conditions le permettent (complexité du réseau, données disponibles, ressources computationnelles), il est favorable d'utiliser un nombre élevé de cartes d'activations pour extraire davantage de caractéristiques.

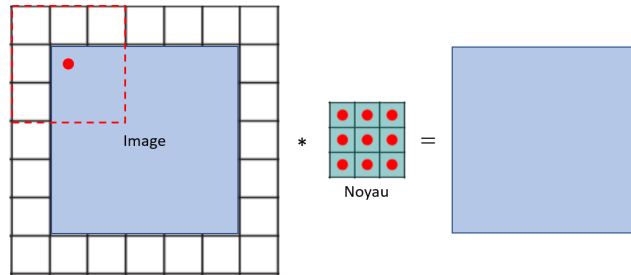


FIGURE 3.8. La convolution d'une image (en bleu) avec un noyau ayant une dimension de $K \times K$ engendre la perte d'une bordure d'une largeur de $\frac{K-1}{2}$ pixels. Avec l'ajout d'une bordure de pixels ayant une intensité nulle (en blanc), la convolution est définie pour tous les pixels de l'image d'origine et résultat est de même dimension que celle-ci.

Pour la convolution, d'autres éléments à considérer incluent l'insertion d'une bordure de P pixels d'intensité nulle (*zero padding*), le pas de déplacement du noyau s et le facteur de dilatation D . L'insertion de zéros est présentée à la figure 3.8. D'abord, effectuer la convolution d'un noyau avec une carte cause la perte d'une région en bordure de l'image, où la convolution n'est pas définie. Dans plusieurs situations, comme lors de l'utilisation de convolutions non dilatées, la région perdue est petite et indésirable. L'insertion d'une bordure de pixels ayant une intensité nulle peut alors être ajoutée pour conserver les dimensions originales. Toutefois, réduire la taille des cartes d'activation peut également être bénéfique. En effet, l'augmentation progressive du champ récepteur est souvent assurée par une réduction contrôlée de la dimension des cartes d'activation. L'insertion de zéros peut alors être omise ou contrôlée de sorte à répondre aux besoins de l'architecture.

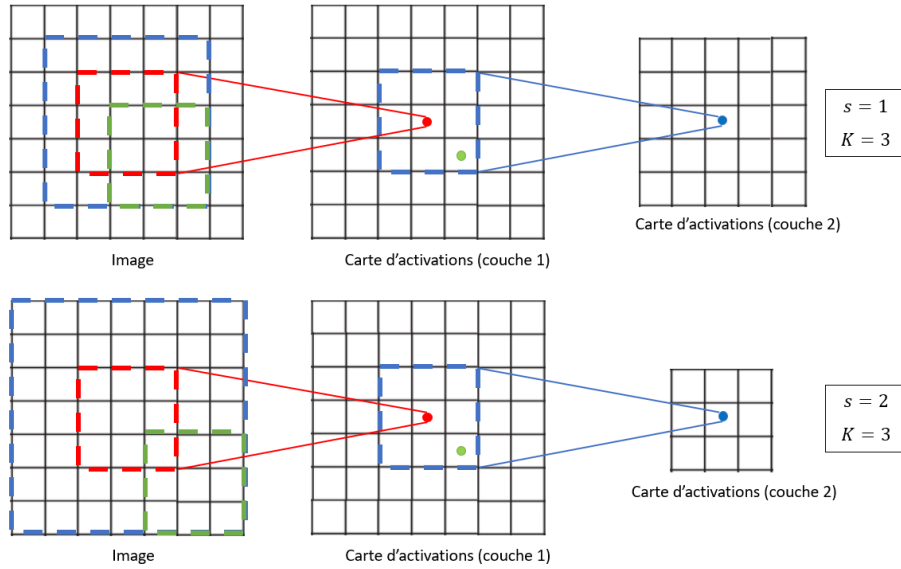


FIGURE 3.9. Comparaison des champs récepteurs d'un neurone de la deuxième couche de convolution (point bleu) avec des noyaux de dimensions $K = 3$ px pour des pas $s = 1$ px (en haut) et $s = 2$ px (en bas). Tous les neurones sont directement connectés à une fenêtre de $K \times K$ neurones de la couche précédente (pointillés de couleur). Toutefois, dans le cas $s = 2$ px, il y a moins de chevauchement entre les champs récepteurs des neurones de la première couche (pointillés rouge et vert) et une plus grande région de l'image est couverte. Le champ récepteur effectif du neurone de la couche 2 est plus élevé (pointillé bleu sur l'image).

Pour ce qui est du pas de convolution et du facteur de dilatation, ces paramètres peuvent être augmentés pour accroître le champ récepteur des neurones. L'augmentation du pas de convolution réduit le chevauchement entre les différentes positions du noyau. Sur la figure 3.9, il est possible de voir l'effet du pas de convolution sur le champ récepteur.

En plus d'augmenter le champ récepteur, accroître le pas permet de réduire la dimension des cartes d'activations. À la sortie d'une couche de convolution, la largeur W'_i d'une carte d'activations est donnée par l'équation suivante [54] :

$$W'_i = \frac{W_i - K_i + 2P_i}{s_i} + 1. \quad (3.4.3)$$

Les convolutions dilatées permettent également d'accroître le champ récepteur du réseau [58]. Aussi appelées convolutions à trous, elles étaient initialement utilisées dans la décomposition en ondelettes [59, 60]. Une convolution dilatée consiste à appliquer une dilatation au noyau avant d'appliquer la convolution. Sommairement, dilater un noyau par un facteur D implique d'insérer un espacement de $D-1$ entre chaque valeur du noyau initial, par l'insertion de zéros dans celui-ci. La dilatation est uniquement faite dans le plan des cartes d'activation. Des exemples de noyaux dilatés sont présentés à la figure 3.10. L'opération de convolution D -dilatée, dénotée $*_D$, entre une carte d'activation et un filtre bidimensionnel est donnée par l'équation 3.4.4 [58].

$$(a_{l-1,i} *_D w_{l,o,i})(\mathbf{p}) = \sum_{\mathbf{s} + D\mathbf{t} = \mathbf{p}} a_{l-1,i}(\mathbf{s}) w_{l,o,i}(\mathbf{t}) \quad (3.4.4)$$

La convolution discrète classique peut être retrouvée avec le facteur de dilatation $D = 1$. Dans les réseaux neuronaux, le principal avantage des convolutions dilatées est qu'elles permettent d'augmenter la taille du champ récepteur. Tel qu'illustré à la figure 3.10, il est possible d'augmenter la taille du champ récepteur de façon exponentielle en augmentant progressivement le facteur de dilatation. De plus, la dilatation du noyau n'ajoute aucun paramètre supplémentaire. Donc, la dilatation peut être utilisée de façon complémentaire avec la succession d'une série de couches de convolutions, sans engendrer de coût supplémentaire. Aussi, contrairement à l'augmentation du pas de convolution et à l'utilisation de pooling (voir la section 3.4.2.2), l'augmentation du champ récepteur ne repose pas sur un sous-échantillonnage des cartes. Ainsi,

un autre avantage des convolutions dilatées est d’augmenter le champ récepteur tout en conservant l’information spatiale. Cette propriété rend les convolutions dilatées particulièrement intéressantes pour la segmentation.

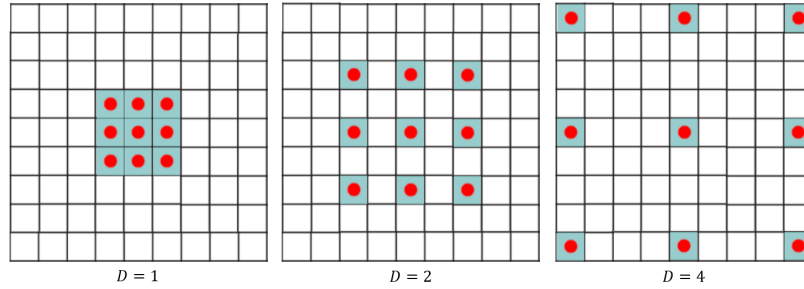


FIGURE 3.10. Un noyau de convolution de dimensions $3 \text{ px} \times 3 \text{ px}$ est dilaté par divers facteurs de dilatation D . Avec $D = 1$, les paramètres du noyau (points en rouge) ne sont pas dilatés. Il est possible de voir qu’en augmentant la dilatation ($D = 2$ au centre, $D = 4$ à droite), le champ récepteur est augmenté. Dans tous les cas, le nombre de paramètres est identique.

Par contre, étant donné la forme des noyaux dilatés (figure 3.10), la dilatation peut introduire des problèmes de maillage (*gridding*), illustrés à la figure 3.11 [28, 61, 62]. En effet, la dilatation consiste à insérer des zéros entre les poids du noyau, donc, bien que le champ récepteur soit plus grand, le noyau couvre en réalité une région analogue à un motif en damier. Cela mène à une perte d’information essentielle sur le voisinage qui s’accroît en augmentant le facteur de dilatation. Toutefois, il est possible d’éviter le problème de maillage en augmentant progressivement le facteur de dilatation utilisé par les couches de convolution successives (figure 3.11) [63–65]. En résumé, l’utilisation de convolutions dilatées avec augmentation progressive des facteurs permet d’augmenter le champ récepteur sans accroître la quantité de paramètres à déterminer durant l’entraînement et sans faire de sous-échantillonnage. Elles ont été utilisées dans plusieurs applications, telles que pour la segmentation des tumeurs cérébrales et de la vessie, de textures pulmonaires, de diverses structures en mammographie, ainsi que dans des applications non médicales comme la segmentation d’images urbaines [28, 66–70].

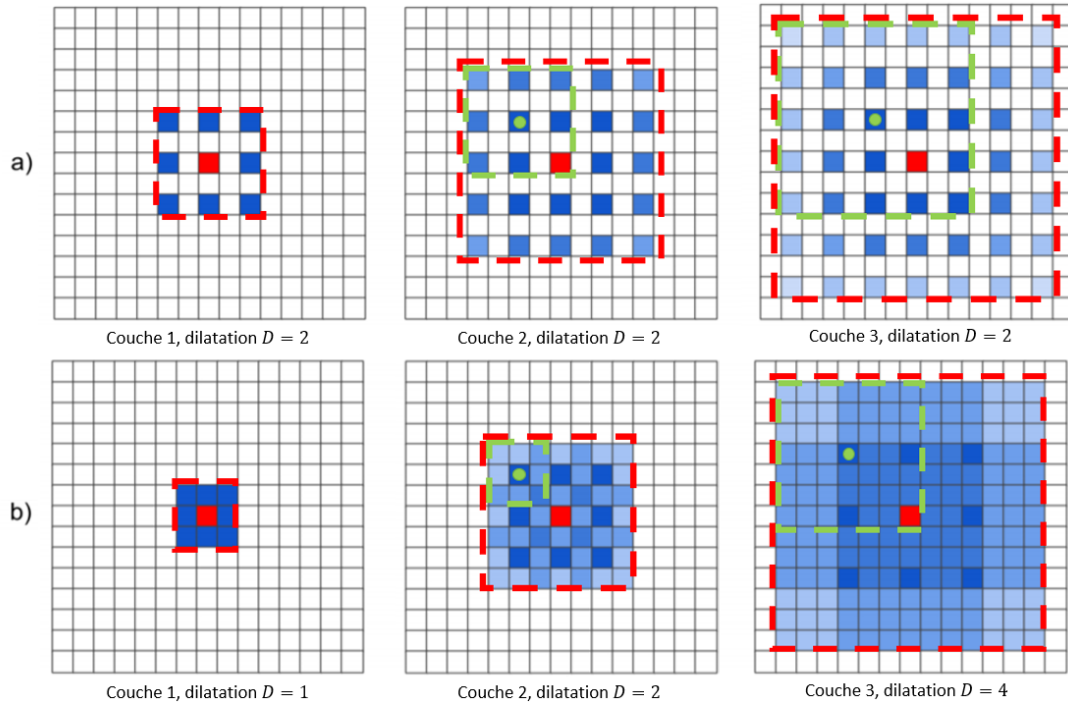


FIGURE 3.11. a) Un problème de maillage survient lorsque le facteur de dilatation n'est pas augmenté progressivement. Par exemple, en dilatant le noyau (carrés bleu foncé) d'un facteur $D = 2$, un neurone de la première couche suivant l'image d'entrée a un champ récepteur de dimensions 3×3 suivant un motif en damier. Le champ récepteur du neurone du centre (carré rouge) est représenté par les pointillés en rouge. À la couche 2, pour calculer l'activation du centre, tous les neurones directement impliqués dans le calcul (par exemple, en vert) dépendent du champ récepteur représenté sur la couche 1. Donc, bien que le champ récepteur soit agrandi à une taille de 7×7 , celui-ci demeure en motif de damier. Les différentes teintes de bleu indiquent le degré de chevauchement entre les champs récepteurs des neurones impliqués dans le calcul. b) En augmentant progressivement le facteur de dilatation, le problème de maillage peut être évité. En utilisant $D = 1$ à la première couche, le noyau original est utilisé. À la couche 2, chaque neurone dépend donc d'une fenêtre 3×3 . Avec cette considération, en appliquant le noyau dilaté ($D = 2$) pour calculer l'activation du neurone rouge, un champ récepteur sans maillage est obtenu. Image adaptée de la référence [65].

3.4.2.2. Pooling

Tel que mentionné dans les sections précédentes, il est commun d'adopter une stratégie pour réduire la dimension des cartes d'activations. D'abord, cela permet d'accroître le champ récepteur des neurones. Aussi, éliminer une partie des activations réduit la quantité de calculs nécessaires, ce qui rend possible l'augmentation progressive du nombre de cartes d'activations. En plus des stratégies introduites, il est commun d'intégrer au réseau des couches de

pooling. Sommairement, le pooling consiste à appliquer un sous-échantillonnage basé sur un critère particulier. Dans un CNN, l'utilisation d'une couche de pooling se fait à la suite d'une couche de convolution.

Le pooling prend en entrée une série de cartes d'activations et un balayage est fait sur celles-ci avec une fenêtre de taille prédéfinie. À chaque position de fenêtre, un critère de sélection est appliqué de sorte qu'une valeur par position de fenêtre soit produite. Les opérations les plus communes sont le max pooling et le pooling moyen avec une fenêtre de $2 \text{ px} \times 2 \text{ px}$, où les valeurs maximales et les valeurs moyennes sont respectivement sélectionnées à chaque position de fenêtre. Avec cette taille de fenêtre, des cartes sous-échantillonnées d'un facteur 2 sont obtenues.

3.4.2.3. *Couche connectée*

Pour faire la classification d'une image, la dernière couche du réseau est une couche connectée, telle que retrouvée dans les perceptrons multicouches. Tel que mentionné précédemment, les couches connectées sont reliées à l'ensemble des neurones de la couche précédente. L'utilisation de ce type de couche permet d'assurer que la sortie du réseau est basée sur l'ensemble des activations. Comme pour les MLP, la dernière couche connectée utilise généralement la fonction softmax comme fonction d'activation et la couche contient un neurone par classe considérée. Lors de l'entraînement, les vecteurs générés par la fonction softmax sont alors envoyés à la fonction coût.

3.4.2.4. *Normalisation de mini-lots*

Plutôt que d'utiliser directement les intensités des voxels comme entrées, les données peuvent être normalisées. Dans le même ordre d'idées, les activations des différentes couches cachées de convolution peuvent également être normalisées. La normalisation de mini-lots (*batch normalization*) est une technique qui contribue à accélérer la convergence de l'algorithme d'apprentissage, à augmenter la stabilité et à améliorer les performances des CNN [71–73].

La normalisation est faite pour chacun des mini-lots, avec l'équation suivante [71] :

$$\zeta_l^{\text{norm}} = \Gamma \frac{\zeta_l - \mathbb{E}(\zeta_l)}{\sqrt{\text{var}(\zeta_l) + \epsilon}} + B, \quad (3.4.5)$$

où ζ_l est la sortie de la couche de convolution l , ϵ est une constante de stabilité numérique, puis Γ et B sont des paramètres d'échelle et de décalage. Les paramètres Γ et B sont appris pendant l'entraînement et sont ensuite utilisés pour la normalisation des données de test. Un des principaux avantages de cette paramétrisation est que l'entraînement par descente de gradient est plus facile et rapide [41]. Un cas particulier de la normalisation de mini-lots est la normalisation d'instance, lorsqu'une seule donnée par inférence est utilisée [28, 74, 75]. La figure 3.12 schématise les dimensions selon lesquelles la normalisation est effectuée avec les normalisations de mini-lots et d'instance. La normalisation d'instance permet une invariance aux changements d'intensité et de contraste. Les caractéristiques extraites sont adaptées pour chacune des coupes, ce qui peut aider, par exemple, lorsque les données utilisées proviennent de plusieurs tomodesitomètres [28].

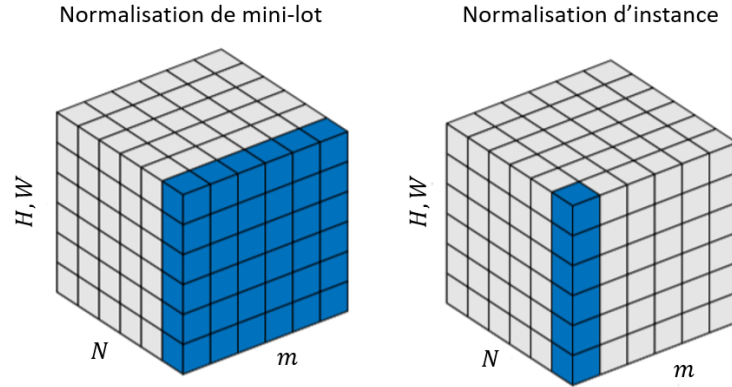


FIGURE 3.12. Les deux cubes correspondent au même tenseur d'activations. La dimension N correspond aux différentes cartes d'activations, m correspond aux données du mini-lot, puis H et W correspondent aux dimensions spatiales. Pour la normalisation de mini-lot, la normalisation est effectuée en utilisant les activations de toutes les données du mini-lot pour chacune des cartes. Pour la normalisation d'instance, la normalisation est faite pour chacune des données. Figure tirée de la référence [76].

3.4.2.5. Connexion raccourci

Les connexions raccourcis (*skip connections*) sont des connexions qui permettent de relier deux couches non adjacentes [77, 78]. Elles peuvent être utilisées pour faire la concaténation de deux ensembles de cartes d’activations (figure 3.13). Ainsi, en reliant deux couches distantes, les connexions permettent d’agréger l’information de bas niveau et spatiale des premières couches avec l’information de haut niveau et sémantique des dernières couches. Dans le même ordre d’idées, l’information résultante d’une connexion raccourci est multiéchelles, puisque les champs récepteurs des deux couches qui combinent leurs cartes diffèrent. Ces connexions sont couramment utilisées dans les problèmes de segmentation, entre autres à l’intérieur des méthodes de récupération de l’information spatiale (voir la section 3.5).

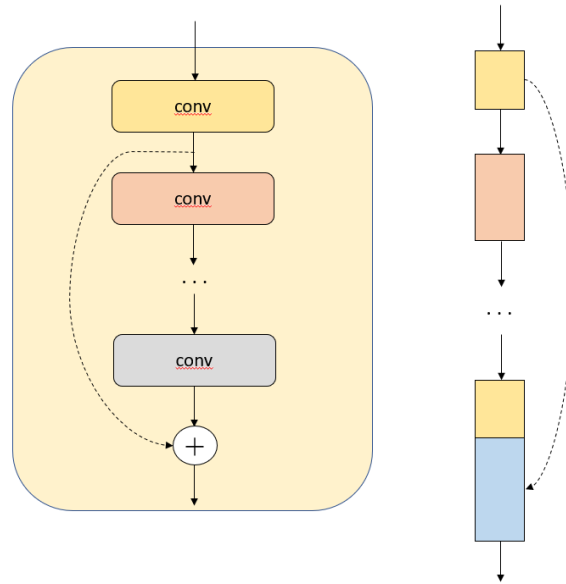


FIGURE 3.13. Deux vues d’une connexion raccourci dans une section de réseau sont présentées. L’information circule de façon séquentielle d’une couche de convolution à l’autre, puis d’autres opérations telles que la normalisation et les fonctions d’activation sont appliquées (représentées par les flèches noires). En plus du chemin séquentiel, une connexion raccourci fait passer directement l’information issue de la première couche de convolution (en jaune) à la dernière (en bleu). Ainsi, les cartes d’activations de ces deux couches sont concaténées, c’est-à-dire qu’à ce point du réseau, le tenseur d’activations contient les cartes d’activations issues des deux couches (illustré à droite).

3.4.2.6. Apprentissage par transfert

Dans le cadre des concours en vision par ordinateur, les réseaux sont entraînés sur des ensembles de données énormes, pouvant atteindre des millions d'images de plusieurs centaines de classes distinctes. Ces réseaux sont généralement assez complexes, puisque la quantité de données disponible permet l'utilisation d'un grand nombre de paramètres. Dans l'optique de tirer profit de ce qui est appris par ces réseaux durant l'entraînement, des approches d'apprentissage par transfert (*transfer learning*) peuvent être considérées. L'apprentissage par transfert vise à utiliser les caractéristiques apprises dans un domaine d'application comme point de départ dans un autre problème. Le principal avantage de l'apprentissage par transfert est que des performances intéressantes peuvent être atteintes avec un nombre limité de données, puisque seul un microajustement (*fine-tuning*) des paramètres est nécessaire.

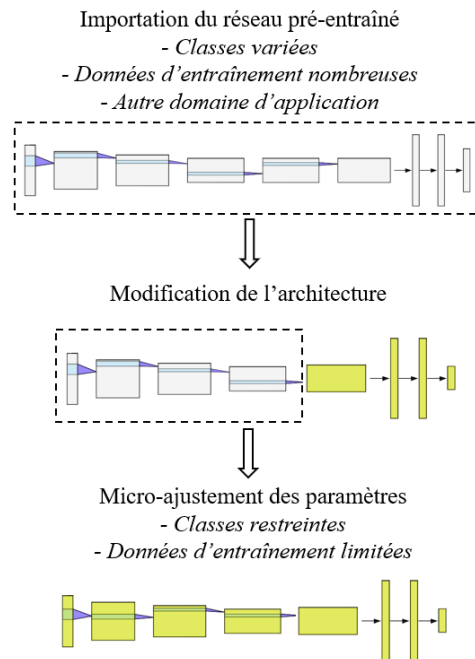


FIGURE 3.14. Les principales étapes permettant d'effectuer un apprentissage par transfert. Le réseau pré-entraîné est d'abord importé et les dernières couches sont remplacées par de nouvelles couches non entraînées. Le réseau résultant peut ensuite être entraîné avec un ensemble de données propre au domaine d'application considéré pour effectuer un microajustement des paramètres.

La figure 3.14 illustre une façon de procéder à un apprentissage par transfert. Les premières couches peuvent extraire des caractéristiques plus générales qui sont applicables au nouveau domaine d'application, tandis que les dernières couches considèrent des caractéristiques plus spécifiques pouvant être inadéquates pour un transfert. Ainsi, une façon de procéder est d'importer le réseau pré-entraîné et de retirer les dernières couches pour les remplacer par des couches qui apprendront des caractéristiques spécifiques au nouveau problème. De plus, comme les premières couches sont déjà entraînées, il est pertinent de modifier les taux d'apprentissage des différentes couches, par exemple, en accélérant l'apprentissage des nouvelles couches ou en stoppant l'entraînement des couches pré-entraînées.

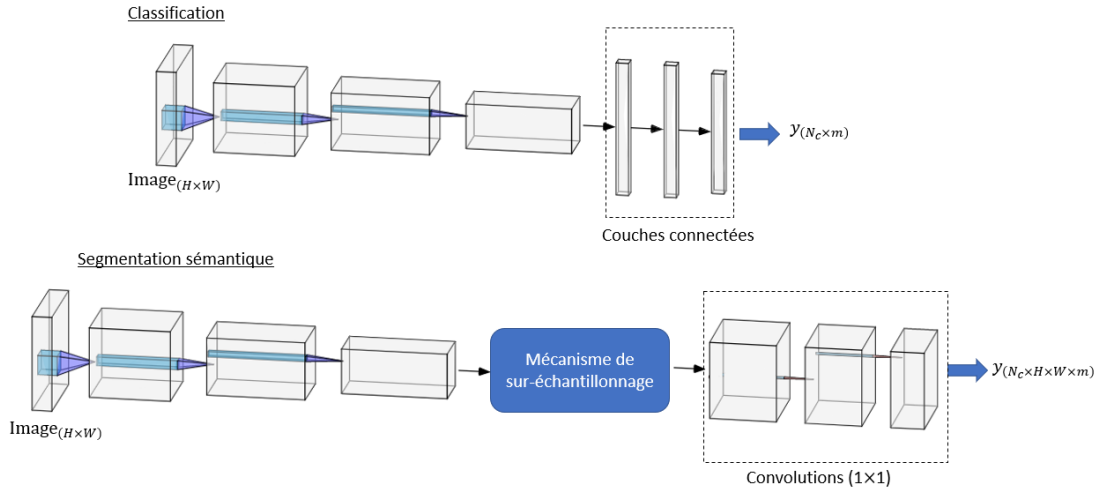


FIGURE 3.15. La différence majeure entre les réseaux de classification et de segmentation est que le réseau de segmentation permet d'assigner une classe à chaque voxel, tandis que le réseau de classification émet une classe par donnée. Donc, par exemple, un réseau de segmentation peut permettre de connaître la localisation des régions atteintes de fibrose dans les poumons, tandis qu'un réseau de classification peut nous informer si les poumons peuvent être qualifiés comme « contenant de la fibrose ». Au niveau de l'architecture, les principales différences sont l'utilisation d'un mécanisme de sur-échantillonnage, qui peut prendre plusieurs formes (couches, connexions raccourcis, etc.), ainsi que le remplacement des couches connectées, qui éliminent l'information spatiale, par des couches de convolutions ayant des noyaux de dimension 1×1 .

3.5. SEGMENTATION SÉMANTIQUE

Dans la section précédente, les CNN ont été présentés. Ceux-ci permettent d’obtenir d’excellentes performances dans la classification d’images, où une étiquette par donnée est émise. Par contre, ces architectures CNN ne permettent pas d’effectuer de segmentation directement, puisque la totalité de l’information spatiale est éliminée par les opérations de sous-échantillonnage et les couches connectées. Dans le cas d’une image CT de fibrose pulmonaire, il serait possible d’extraire des petites régions d’intérêt sur l’image et de les utiliser avec un CNN pour faire la classification des régions. Par contre, en ramenant les régions d’intérêt sur l’image d’origine, la segmentation serait relativement grossière et approximative.

Contrairement à la classification, la segmentation sémantique vise à attribuer une classe par voxel plutôt qu’une classe par donnée. La distinction entre une classification et une segmentation par réseau de neurones est illustrée à la figure 3.15. En classification, la sortie du réseau consiste en un vecteur de N_c valeurs (une étiquette par donnée). En segmentation sémantique, la sortie correspond plutôt à un volume de dimensions $N_c \times H \times W$ pour chaque donnée, ce qui correspond à une carte de segmentation par classe.

Les opérations de sous-échantillonnage ont pour effet d’éliminer une partie de l’information spatiale. Comme l’objectif en segmentation est d’obtenir des cartes de segmentation ayant les mêmes dimensions spatiales que la donnée d’entrée, un mécanisme de sur-échantillonnage (*upsampling*) est introduit afin de récupérer progressivement l’information spatiale nécessaire, tout en gardant l’information sémantique. En d’autres mots, lorsqu’elle est appliquée à une carte d’activations, une opération de sur-échantillonnage consiste à augmenter le nombre de lignes et de colonnes de celle-ci, généralement d’un facteur 2. Des exemples de techniques pouvant être utilisées pour le sur-échantillonnage incluent l’interpolation bilinéaire et les convolutions transposées. Dans ce contexte, la transposition fait référence au fait que la multiplicité de l’opération (1 entrée à plusieurs sorties) est l’inverse de celle d’une convolution standard (plusieurs entrées à 1 sortie) [79]. Par exemple, si une convolution diminue les dimensions de l’entrée d’un facteur 2, alors sa transposée augmentera les

dimensions du même facteur. En réalité, les opérations effectuées sont des convolutions apprises durant l’entraînement qui utilisent une insertion de zéros adéquate pour assurer le sur-échantillonnage. La figure 3.16 présente un exemple de convolution transposée pouvant être intégrée dans un CNN. Afin d’utiliser

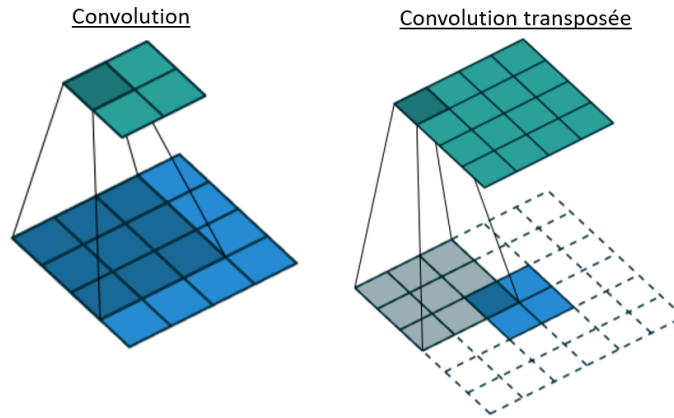


FIGURE 3.16. Opération de convolution (à gauche) entre une entrée (en bleu) de dimension 4×4 et un noyau (en bleu foncé) de dimension 3×3 ainsi que la convolution transposée (à droite). La convolution donne un résultat (en turquoise) de dimension 2×2 , alors que sa transposée permet d’inverser cette multiplicité en passant d’une entrée 2×2 à un résultat 4×4 . Pour ce faire, une convolution standard peut être appliquée en insérant une bordure de zéros adéquate ($P=2$). Dans un CNN, les entrées et les sorties seraient les cartes d’activations, puis le noyau serait appris durant l’entraînement. Figure adaptée de la référence [79].

l’information spatiale des premières couches dans les opérations de sur-échantillonnage, des connexions raccourcis peuvent être utilisées entre les premières couches et les couches de sur-échantillonnage. Généralement, les cartes d’activations des deux couches sont additionnées ou concaténées [80, 81]. Lorsque des opérations de pooling sont utilisées, il est également possible d’envoyer à la couche de sur-échantillonnage les coordonnées originales des activations conservées par l’opération [82]. En plus du sur-échantillonnage, l’autre aspect à considérer est le remplacement des couches connectées. En effet, les couches connectées retrouvées à la fin du réseau éliminent la totalité de l’information spatiale restante pour produire un résultat purement sémantique.

Ces couches connectées peuvent être vues comme des convolutions avec un noyau qui a les dimensions du volume d'entrée. Afin d'obtenir une prédiction par voxel, celles-ci sont remplacées par des couches de convolution qui utilisent des noyaux de dimension 1×1 [80]. L'utilisation de ces couches permet de réduire progressivement le nombre de cartes d'activations, sans réduire les dimensions spatiales de celles-ci. Ainsi, combiné avec le sur-échantillonnage, toute l'information sémantique présente dans les cartes d'activation est finalement combinée pour produire des cartes d'activations qui ont les mêmes dimensions que l'image originale (une prédiction par voxel). Pour avoir les cartes de segmentation finales, la dernière couche de convolution 1×1 produit une carte d'activations par classe considérée. Comme ces réseaux ne contiennent plus de couches connectées, ces CNN sont aussi référés comme étant des réseaux de neurones complètement convolutifs (*fully convolutional neural network*).

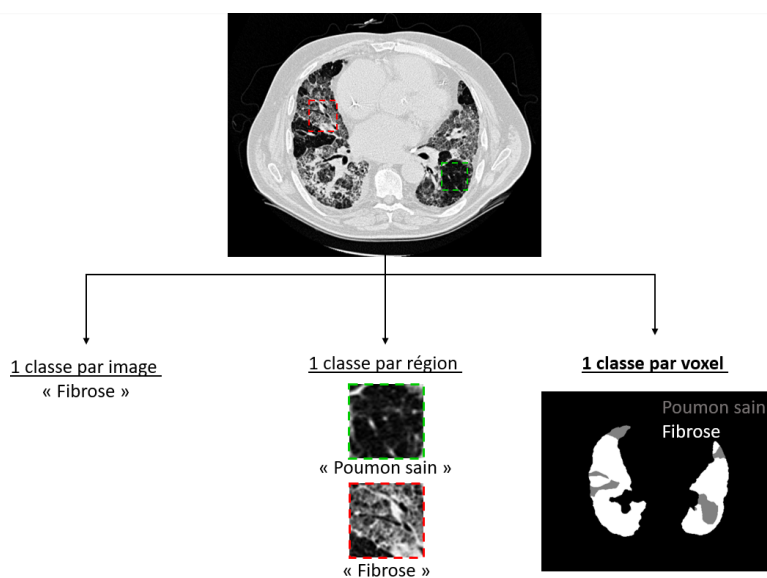


FIGURE 3.17. Pour la phase d'entraînement, les réseaux de segmentation requièrent généralement des annotations beaucoup plus difficiles à produire qu'en classification. Pour une classification d'images, il serait possible d'assigner la classe Fibrose à l'image, qui semble majoritairement contenir ce type de texture pulmonaire. Pour une classification de régions d'intérêt (pointillées de couleur), il faudrait plutôt assigner une classe à chacune des régions extraites sur l'image. En segmentation, pour chaque image utilisée, un masque de segmentation est produit (voxels sains en gris, voxels de fibrose en blanc).

Les problèmes de segmentation ont généralement des difficultés additionnelles, comparativement à la classification, notamment à cause de la quantité de prédictions faites pour chaque image. Pour cette raison, alors que des problèmes de classification peuvent considérer des centaines de classes distinctes, la segmentation se limite généralement à un nombre beaucoup plus restreint de classes. De plus, obtenir les classes réelles (réalités de terrain, *ground truth*) nécessaires à l'apprentissage supervisé est beaucoup plus coûteux que pour les réseaux de classification, ce qui limite grandement la quantité de données disponibles. Les différents types d'annotations sont présentés à la figure 3.17.

Dans la plupart des cas, les réalités de terrain fournies au réseau doivent être faites à la même échelle que les prédictions qui seront faites par le réseau. Ainsi, pour une classification d'images, il serait nécessaire de spécifier à quelle classe correspond l'image globalement (une étiquette par image), tandis que pour une classification de régions d'intérêts, une classe par région devrait être spécifiée. Du côté de la segmentation, il faut assigner une classe à une grande quantité de voxels pour chacune des images. Dans la prochaine section, le CNN U-Net est introduit [81].

3.5.1. Architecture U-Net

Le U-Net est l'architecture gagnante de l'édition 2015 d'un concours de l'*IEEE Symposium on Biomedical Imaging*, où la segmentation sémantique d'images contenant des cellules devait être effectuée dans un contexte où la quantité de données est limitée [81]. L'excellente performance du U-Net a mené à de nombreuses variantes de l'architecture utilisées dans plusieurs applications médicales. Des exemples d'applications médicales incluent la segmentation de structures et de lésions dans le foie, la vessie, le cerveau, le colon et le coeur [65, 83–86]. Les détails de l'architecture sont présentés à la figure 3.18. D'abord, le CNN contient un encodeur (chemin de contraction, *contracting path*) suivi d'un décodeur (chemin d'expansion, *expansive path*). Le rôle de l'encodeur est celui d'un CNN standard, c'est-à-dire transformer l'image d'entrée en une série de cartes d'activations qui contiennent de l'information sémantique de haut niveau.

Du côté encodeur, la sortie d'un module se fait par un pooling maximal 2×2 ainsi qu'une augmentation d'un facteur 2 du nombre de cartes, de sorte à garder une quantité similaire de neurones par couche. Du côté décodeur, les différents modules sont plutôt connectés par un sur-échantillonnage d'un facteur 2, ainsi qu'une diminution d'un facteur 2 du nombre de cartes. Des connexions raccourcis sont présentes entre les couches encodeur et décodeur de même niveau, où les cartes sont de mêmes dimensions spatiales. Cela permet de combiner, par concaténation, différents niveaux d'abstraction dans la partie décodeur. Par exemple, une connexion raccourci est présente entre le dernier bloc du module encodeur sous-échantillonné 1 fois et le premier bloc du module décodeur sur-échantillonné 3 fois. Finalement, une couche de convolution 1×1 est utilisée afin de réduire la quantité de cartes à une par classe considérée. Dans l'implémentation originale, aucune insertion de zéros n'est utilisé, ce qui explique la différence entre les dimensions spatiales en entrée et en sortie. En insérant une bordure de zéros, il est possible de ne pas engendrer une perte de bordure et de conserver les dimensions originales.

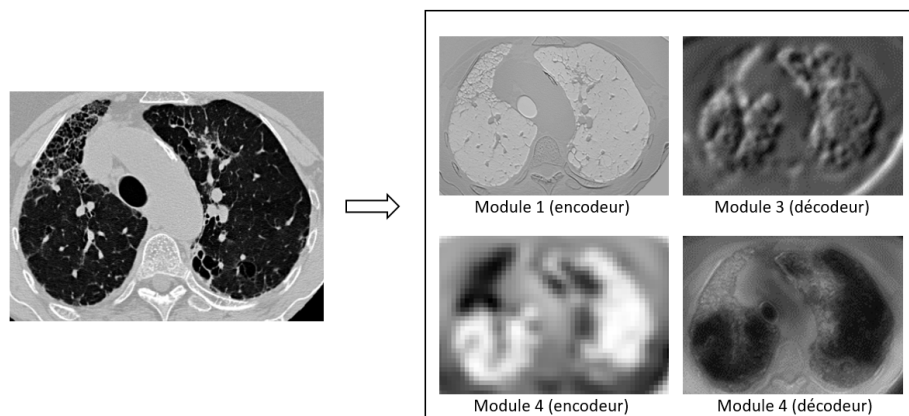


FIGURE 3.19. Des exemples de cartes d'activations suivant une inférence du U-Net sur l'image de gauche. Dans le premier module de l'encodeur, les cartes d'activations représentent majoritairement de l'information spatiale de bas niveau, tel que la présence de certains contours. Dans le quatrième module de l'encodeur, la résolution spatiale des cartes est fortement dégradée (sous-échantillonnage d'un facteur 8). Les cartes représentent des abstractions de haut niveau comme la présence de certaines textures. Du côté décodeur, de l'information spatiale est graduellement récupérée tout en conservant l'information sémantique extraite par l'encodeur.

Chapitre 4

MÉTHODOLOGIE

Ce chapitre présente la méthodologie utilisée pour atteindre les objectifs du projet. Pour effectuer la segmentation, la méthodologie est basée sur les CNN complètement convolutifs, abordés lors de la section 3.5. Ceux-ci permettent d'obtenir des prédictions denses sur la présence des différents motifs en une seule inférence. Les CNN complètement convolutifs considérés sont utilisés dans des contextes d'apprentissage supervisé et semi-supervisé. Dans les deux scénarios, des images contenant des annotations effectuées par des experts sont nécessaires. Ces données sont utilisées pour la phase d'entraînement des CNN et pour l'évaluation des performances.

Dans les prochaines sections de ce chapitre, les données des patients et les architectures de CNN qui ont été utilisées au cours du projet sont abordées. Par la suite, la méthodologie en lien avec l'entraînement, la validation et le test des réseaux est discutée. Les différentes métriques considérées pour l'évaluation quantitative des performances sont également introduites.

4.1. IMAGES DES PATIENTS ET ANNOTATIONS

Dans cette section, plusieurs aspects concernant les images patients et les annotations sont abordés. D'abord, une description des données nécessaires pour atteindre les objectifs du projet est faite. Par la suite, l'ensemble de données disponible qui a été utilisé au cours du projet est discuté.

4.1.1. Description des données requises

L'optimisation des paramètres des CNN et l'évaluation de leurs performances requièrent un ensemble de données. Cet ensemble doit contenir une série d'images ainsi que les réalités de terrain (*ground truth*) associées à celles-ci. Dans le contexte de ce projet, les réalités de terrain consistent en des annotations de référence. La série d'images correspond à des images CT pulmonaires de plusieurs patients souffrant de maladies pulmonaires et les

annotations consistent en des délinéations effectuées sur les images par des experts. Ces délinéations permettent d’assigner une classe aux différents voxels qui servent lors de l’entraînement et pour vérifier les prédictions du réseau dans un contexte de segmentation sémantique. Concrètement, les annotations peuvent être présentées sous forme de masques, où l’intensité des pixels est utilisée pour attribuer la classe. Par exemple, un masque binaire pourrait être généré en assignant une intensité de 255 aux voxels sains et une intensité nulle aux voxels de fibrose. Les annotations peuvent aussi être représentées sous forme de contours à l’intérieur desquels les voxels se voient attribuer la classe associée au contour.

Pour segmenter la fibrose pulmonaire, plusieurs motifs distincts doivent être considérés, notamment le poumon sain, le verre dépoli, les motifs en rayons de miel, la réticulation, la bronchiectasie de traction et les distorsions d’architecture [6]. Avec l’objectif de fournir de l’information sur la présence de motifs associés au diagnostic histologique de fibrose pulmonaire, ces textures pulmonaires peuvent être regroupés sous trois classes distinctes, résumées dans le tableau 4.1. D’abord, la classe Sain correspond aux tissus pulmonaires qui ne

TABLEAU 4.1. Textures pulmonaires incluses dans chacune des classes considérées pour la segmentation

Classes	Motifs
Sain	Tissus sans motifs d’ILD
Verre dépoli	Tous les degrés de verre dépoli
Fibrose	Rayons de miel
	Réticulation
	Bronchiectasie de traction
	Distorsions d’architecture

sont pas associés à une ILD. Ensuite, la classe Verre dépoli correspond à des motifs contenant différents degrés d’épaississement des tissus pulmonaires. Cette classe n’est pas spécifique à une ILD. La classe Fibrose correspond plutôt aux motifs qui peuvent être associés au diagnostic de fibrose pulmonaire et inclut les motifs de rayons de miel, la réticulation, la bronchiectasie de traction et les distorsions d’architecture. Des régions d’intérêts contenant des tissus de chacune des classes sont présentées à la figure 4.1. Dans la prochaine section, l’ensemble de données utilisé au cours du projet est introduit.

4.1.2. Ensemble de données

Les données utilisées au cours de ce projet proviennent d'un ensemble de données fourni par les Hôpitaux universitaires de Genève (HUG). L'ensemble utilisé fait partie d'une base de données publique d'images HRCT pulmonaires qui contient les images et les annotations de 108 patients souffrant d'ILD [6]. Les données incluent 13 ILD différentes, dont notamment la fibrose pulmonaire. Au niveau des annotations, la délinéation des différents motifs a été faite par deux radiologistes ayant 15 et 20 ans d'expérience.

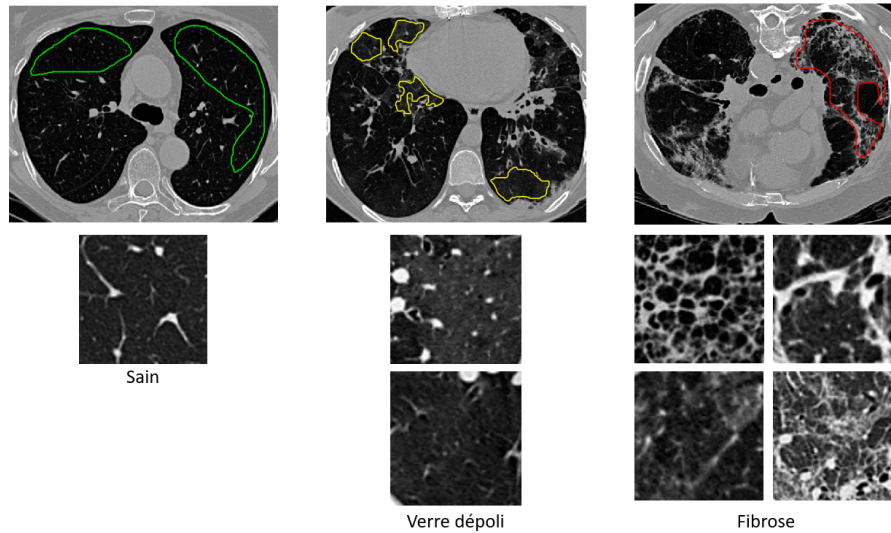


FIGURE 4.1. Des exemples d'annotations et de régions d'intérêt sont présentées pour les 3 types de tissus considérés. Sur les images annotées, tous les voxels présents à l'intérieur des contours de couleur correspondent à la classe représentée par le contour (poumon sain en vert, verre dépoli en jaune et fibrose en rouge).

Afin de réduire l'effort requis par patient, uniquement des annotations partielles ont été effectuées. Pour ce faire, pour chacun des patients, un sous-ensemble des coupes est sélectionné et uniquement quelques régions sont annotées sur celles-ci. Les délinéations faites par les radiologistes couvrent les régions où ceux-ci étaient confiants en leurs prédictions. Cela signifie que, par exemple, si une image contient un contour de fibrose, il peut également y avoir de la fibrose non annotée à l'extérieur du contour. La sélection des coupes est faite avec un pas fixe afin de couvrir la totalité du patient. Cela a permis d'annoter plus de

patients, ce qui a l'avantage d'augmenter la diversité des motifs, qui peuvent varier de façon importante d'un patient à l'autre. Par contre, les annotations partielles apportent quelques difficultés. Notamment, les données disponibles ne sont pas volumétriques et pour chacune des coupes, seulement des réalités de terrain partielles sont disponibles, sous forme de masques partiels. Puisque les voxels non annotés n'ont pas de classe réelle pour évaluer la fonction coût, ceux-ci ne contribuent pas à l'apprentissage supervisé et à l'évaluation des prédictions. En plus des masques partiels, pour chacune des coupes, un masque de poumon entier est disponible. Celui-ci peut être utilisé pour restreindre la segmentation à la région pulmonaire. La figure 4.2 présente un exemple d'une donnée disponible dans la base de données (l'image et les deux masques associés).

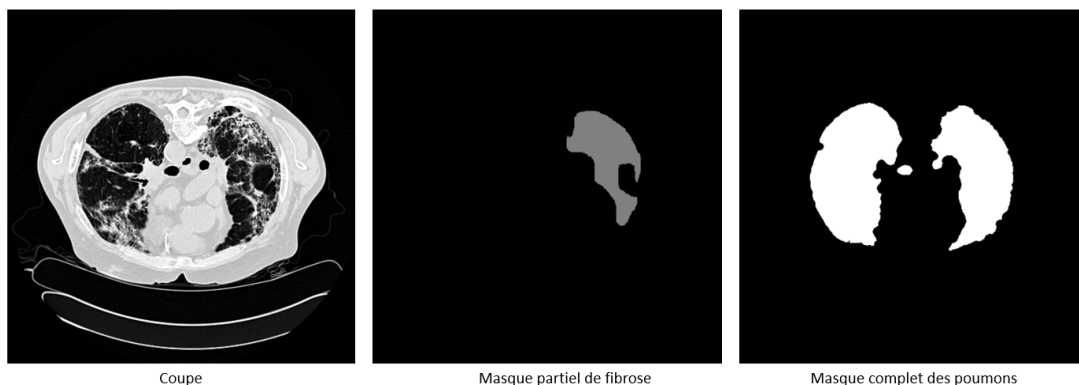


FIGURE 4.2. Toutes les données disponibles consistent en une image (à gauche), un masque partiel d'annotations (au centre) et un masque complet de poumons (à droite). Les régions annotées (en gris, au centre) ne sont pas complètes, donc il peut aussi y avoir des motifs non annotés à l'extérieur du masque.

Pour répondre aux objectifs du projet, les images contenant des annotations de tissus sains, de verre dépoli et de fibrose ont été sélectionnées. Pour l'entraînement et la validation des CNN considérés, 545 images provenant de 64 patients ont été utilisées. Le tableau 4.2 résume les principaux paramètres d'images et les détails sur la quantité de données disponibles. Avec ces données, l'entraînement du réseau peut être effectué et ses performances peuvent être évaluées sur des données du même domaine (sur des annotations partielles). Ces données permettent d'évaluer le réseau sur une grande diversité de tissus. Les performances ont également été évaluées sur des annotations complètes réalisées

au CHUM par 3 radio-oncologues. Cela permet de considérer un scénario plus près des applications cliniques envisagées sur un nombre restreint de patients. Ainsi, en plus des images de validation, des images de test ayant des réalités de terrain complètes sont utilisées.

TABLEAU 4.2. Paramètres d’images et statistiques des données d’entraînement et de validation pour les classes Sain (S), Verre dépoli (V) et Fibrose (F).

Espacement (x,y)	0.42 - 0.98 mm
Champ de vue (x,y)	278 - 499 mm
Épaisseur des coupes	1 mm
Espacement des coupes	10 mm
Images annotées (S, V, F)	54, 242, 254
Voxels annotés ($\times 10^5$) (S, V, F)	7.38, 10.6, 11.6

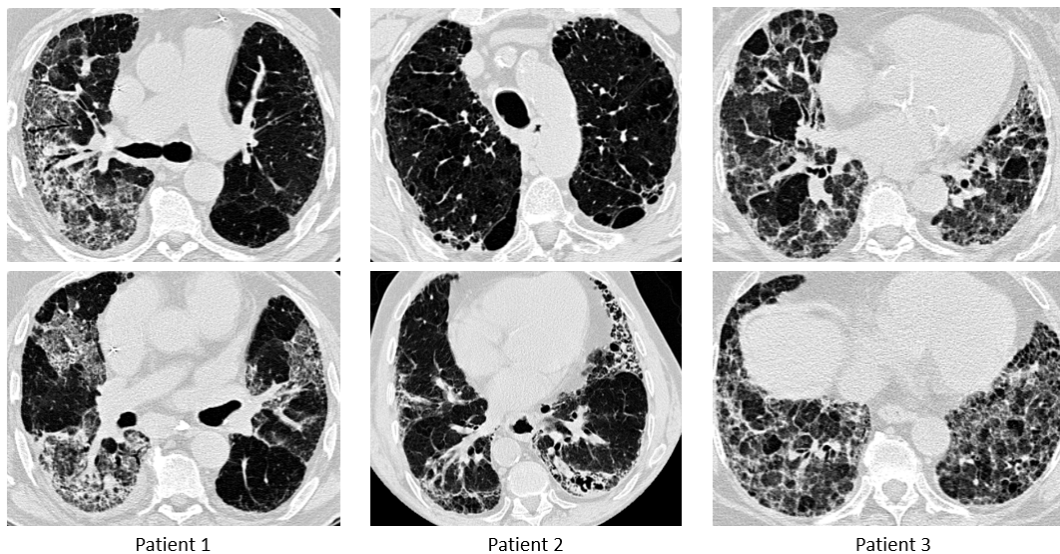


FIGURE 4.3. Des exemples d’images des patients utilisés pour les tests. Les patients 1 et 2 (à gauche et au centre) contiennent des régions de fibrose généralement plus petites ou subtiles, tandis que le patient 3 (à droite) contient des régions de fibrose plus sévères.

Les annotations complètes ont été réalisées sur 3 patients sélectionnés dans la base de données des HUG, afin d’être utilisés comme patients de test exclus de l’entraînement. La sélection a été effectuée de sorte à avoir uniquement des patients ayant reçu un diagnostic de fibrose pulmonaire et peu de motifs des

autres classes non considérées, tels que l’emphysème. Le degré de sévérité et l’étendue des motifs de fibrose sont très différents chez les 3 patients. La figure 4.3 présente des images de chacun des patients de test. Les annotations ont été effectuées pour 2 classes, soit la classe Sain et la classe Risque Fibrose Pulmonaire (RFP), qui combine tous les motifs pouvant être associés à la fibrose pulmonaire (le verre dépoli et les motifs de fibrose). Les statistiques des données de test sont présentées dans le tableau 4.3. En résumé, pour l’optimisation des paramètres, des annotations partielles sont utilisées.

TABLEAU 4.3. Images de test disponibles pour chaque classe (Sain (S) et Risque Fibrose Pulmonaire (RFP)) et au total, ainsi que le nombre de voxels pour lesquels une réalité de terrain (GT) est disponible

Patient	Images _S	Images _{RFP}	Images _{Total}	GT _S ($\times 10^5$)	GT _{RFP} ($\times 10^5$)
Patient 1	20	21	21	2.04	3.16
Patient 2	18	19	19	5.15	3.67
Patient 3	13	20	20	0.25	3.57

4.2. ÉVALUATION DES PERFORMANCES

En plus de l’entraînement, les données présentées dans la dernière section sont utilisées pour évaluer les performances des CNN à l’aide des métriques introduites dans cette section.

4.2.1. Matrice de confusion

D’abord, des matrices de confusion telles que présentées à la figure 4.4 sont utilisées. Les valeurs d’une matrice de confusion représentent le nombre d’assignations qui ont été faites pour chaque assignation possible. Les assignations possibles correspondent aux relations entre les différentes classes prédites et réelles. Par exemple, en considérant la première colonne, les entrées (1,1), (2,1) et (3,1) correspondent aux nombres de voxels qui se sont fait assigner la classe C_1 alors que les classes réelles étaient C_1 , C_2 et C_3 respectivement. Une ligne représente les assignations des classes C_1 , C_2 et C_3 pour une certaine classe réelle. Pour un problème d’assignation à N_c classes, la matrice de confusion est de dimension $N_c \times N_c$. Les quatre types d’assignations (figure 4.5), soit les vrais positifs (VP), vrais négatifs (VN), faux positifs (FP)

et faux négatifs (FN), peuvent être déterminés à partir de la matrice de confusion (figure 4.4). À partir de la matrice de confusion, plusieurs métriques peuvent être déterminées. La prochaine section discute des métriques qui sont dérivées de la matrice de confusion pour évaluer les performances.

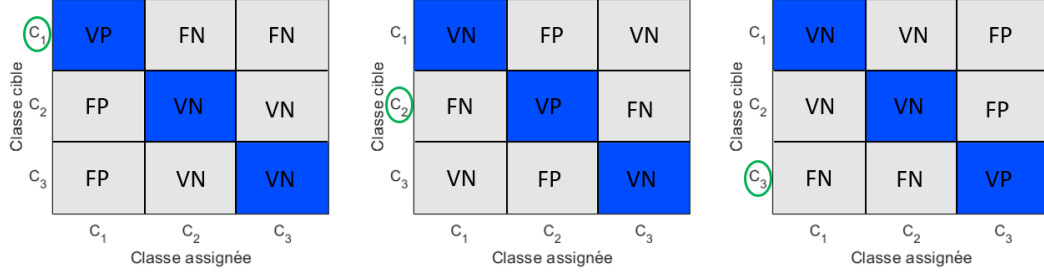


FIGURE 4.4. Les 3 matrices présentées correspondent à la même matrice de confusion à 3 classes. Dans chaque matrice, les différents types d'assignations par rapport aux classes C_1 (à gauche), C_2 (au centre) et C_3 (à droite) sont indiquées. Par exemple, par rapport à la classe C_1 , les assignations qui sont des classes C_2 et C_3 , alors que la classe réelle est C_1 (entrées (1,2) et (1,3)), correspondent à des faux négatifs (FN). Inversement, les assignations de classe C_1 alors que les classes réelles étaient C_2 et C_3 (entrées (2,1) et (3,1)) correspondent à des faux positifs (FP) de la classe C_1 .

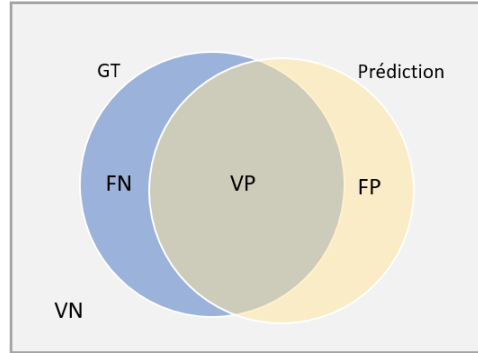


FIGURE 4.5. En comparant une prédiction du réseau avec l'annotation de référence (GT) d'une classe, 4 types d'assignations sont possibles. Par exemple, si la classe Fibrose était assignée et que c'est également la classe réelle du voxel, l'assignation serait un vrai positif (VP), tandis qu'elle serait un faux négatif (FN) par rapport à la fibrose si le réseau avait prédit une autre classe. Inversement, si le voxel correspond à des tissus sains et que de la fibrose lui était assignée, ce serait un faux positif par rapport à la classe fibrose.

4.2.2. Métriques dérivées

Une première métrique calculée est la justesse (*accuracy*) d'assignation A . Celle-ci permet d'évaluer globalement la capacité du réseau à assigner la bonne classe en calculant la fraction des voxels qui ont été assignés au bon type de tissu. À partir d'une matrice de confusion M , la justesse d'assignation est donnée par l'équation suivante :

$$A = \frac{\text{Tr}(M)}{\sum_{i,j=1}^{N_c} M(i,j)}. \quad (4.2.1)$$

Le recouvrement (*recall*) Re est également évalué pour chacune des classes individuellement. Celle-ci nous informe sur la capacité du réseau à détecter les différents types de tissus. Pour une classe i , le recouvrement est donnée par l'équation suivante :

$$Re_i = \frac{M(i,i)}{\sum_{j=1}^{N_c} M(i,j)}, \quad (4.2.2)$$

et, en termes des types d'assignations, $Re_i = \frac{TP_i}{TP_i + FN_i}$. La précision Pr est également évaluée et nous indique la tendance avec laquelle le réseau attribue type de tissu à des voxels qui sont en réalité d'une autre classe. Cette métrique est complémentaire au recouvrement, qui n'inclut pas les faux négatifs. La précision permet, entre autres, de détecter des situations où le recouvrement pour un type de tissu est élevé à cause d'une prédiction abusive de ce type de tissu à chacun des voxels. Pour une classe i , la précision est donnée par l'équation suivante :

$$Pr_i = \frac{M(i,i)}{\sum_{j=1}^{N_c} M(j,i)}, \quad (4.2.3)$$

et, en termes des types d'assignations, $Pr_i = \frac{TP_i}{TP_i + FP_i}$.

La dernière métrique considérée est le coefficient Dice (DSC) [87]. Le DSC est parmi les métriques les plus utilisées en segmentation d'images médicales [88]. Celui-ci permet d'évaluer la similarité entre une carte de segmentation prédite CS_p et la réalité de terrain CS_{GT} . Pour une classe i , le coefficient Dice est donné par l'équation suivante :

$$DSC_i = \frac{2|CS_{p,i} \cap CS_{GT,i}|}{|CS_{p,i}| + |CS_{GT,i}|}. \quad (4.2.4)$$

Au numérateur, la cardinalité de l'intersection mesure les vrais positifs, tandis que le dénominateur pénalise autant les faux positifs que les faux négatifs. Le

DSC convient particulièrement bien aux problèmes où tenir compte des deux types d'erreurs est essentiel. En termes des types d'assignations, le coefficient est donné par l'équation 4.2.5.

$$DSC_i = \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (4.2.5)$$

Tel qu'il est possible de le voir avec l'équation 4.2.5, le DSC peut s'interpréter comme une combinaison de la spécificité et de la précision. Un DSC élevé (près de 1) peut s'interpréter comme une segmentation prédite qui concorde bien avec les annotations. Dans le cas d'un DSC plus faible, la précision et le recouvrement peuvent nous informer davantage sur le type d'assignation qui limite les performances du modèle.

4.3. LUNGNET

Au cours des prochaines sections, les architectures implémentées sont discutées. Ensuite, les détails liés à l'entraînement des CNN sont abordés. Dans un premier temps, l'architecture utilisée dans la méthode LungNet correspond au CNN complètement convolutif proposé par Anthimopoulos *et al* dans la segmentation de certaines pathologies pulmonaires telles que les micronodules [28]. Celle-ci a été implémentée et entraînée dans le cadre de ce projet afin d'effectuer la segmentation des 3 types de tissus considérés. L'architecture est d'abord discutée, suivie des principaux détails d'implémentation.

4.3.1. Architecture

LungNet est un CNN totalement convolutif, c'est-à-dire qu'il ne contient aucune couche connectée. Un schéma de l'architecture est présenté à la figure 4.6. La sortie du réseau consiste en 3 cartes d'activations (une par classe) organisées en un volume ayant les mêmes dimensions spatiales que l'entrée (512×512). L'augmentation du champ récepteur repose sur l'utilisation de convolutions dilatées avec des facteurs de dilatation augmentés progressivement. Le CNN contient un total de 10 couches de convolutions dilatées et les facteurs de dilatation utilisés correspondent aux 10 premiers termes de la suite de Fibonacci (1, 1, 2, 3, 5, 8, 13, 21, 34 et 55), de sorte à diminuer le problème de maillage comparativement à une augmentation exponentielle ($D = 1$ à $D = 64$). Les dimensions spatiales des cartes demeurent constantes par l'utilisation d'une

insertion de zéros. À la suite de chaque convolution, une normalisation d'instance est effectuée, puis une combinaison des activations normalisées et non normalisées est effectuée avant d'appliquer une fonction d'activation ReLU. Chacune des 10 couches utilise 32 cartes d'activations et, à la fin du réseau, la concaténation de celles-ci avec l'image d'entrée est effectuée, afin d'avoir un volume contenant 321 cartes d'activations. Cette concaténation permet de combiner l'information provenant de plusieurs niveaux d'abstraction et échelles. Lors de l'entraînement, l'abandon est utilisé comme technique de régularisation. Finalement, le volume passe par une série de convolutions 1×1 pour réduire progressivement le nombre de cartes à une par classe et la sortie est convertie en distribution de probabilités par la fonction softmax.

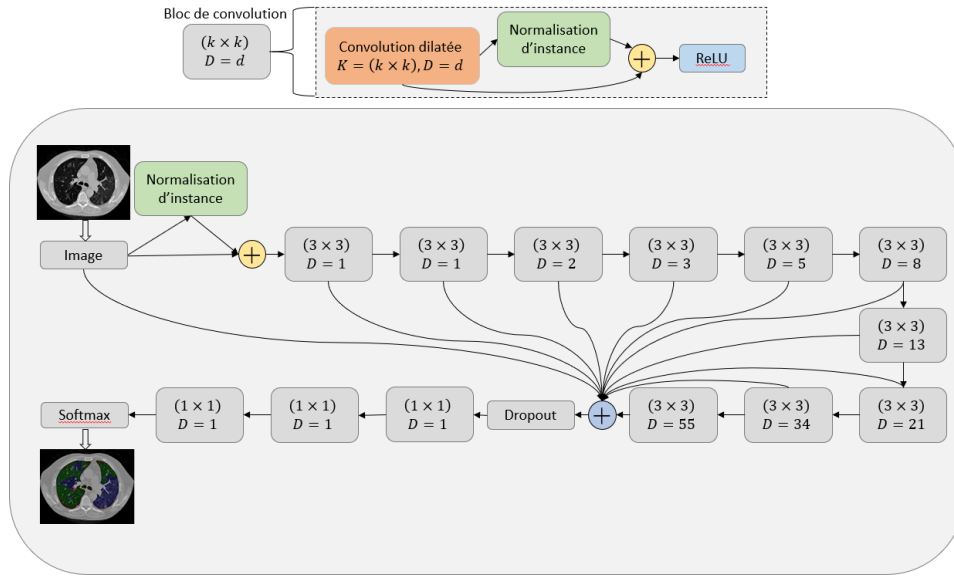


FIGURE 4.6. L'architecture LungNet utilise une série de blocs de convolutions dilatées pour effectuer des prédictions denses (une par voxel) en une seule inférence. En haut, les composantes du bloc de convolution sont présentées, soit une couche de convolution avec une taille de noyau K et un facteur de dilatation D (en rouge), une normalisation d'instance (en vert) et l'application d'une unité rectifiée linéaire (en bleu). Les activations avec et sans normalisation sont sommées (en jaune) à l'aide de la connexion raccourci. En bas, l'architecture basée sur ce bloc (en gris) est présentée. À la suite des 10 blocs de convolutions, la concaténation (en bleu) de l'ensemble des cartes d'activations est effectuée.

4.3.2. Détails d'implémentation

L'entraînement et la validation ont été effectués suivant la méthode proposée dans la méthode LungNet. L'entraînement est effectué jusqu'à ce qu'aucune amélioration de la justesse d'assignation ne se fasse sur les données de validation pendant 50 époques. Comme la quantité de données est relativement limitée, une validation croisée à 5 volets a été effectuée. Pour ce faire, les données disponibles ont d'abord été séparées en 5 ensembles. Ensuite, pour un volet, l'entraînement et la validation sont effectués respectivement sur 4 ensembles et 1 ensemble. Cette procédure est répétée pour chacun des 5 volets en permutant l'ensemble de validation. Les données ont été séparées par rapport aux patients, de sorte que les tissus d'un patient ne se retrouvent pas dans plus d'un ensemble. Un débalancement abusif des données pourrait survenir si un type de tissu est sous-représenté dans l'ensemble d'entraînement d'un volet. Pour éviter cette situation, l'entropie de la distribution de classe des 5 volets est maximisée itérativement. À partir d'une configuration de départ quelconque, une itération consiste à échanger deux patients de deux ensembles. Si l'entropie est supérieure à l'itération précédente, cette nouvelle configuration sera utilisée pour l'itération suivante. L'implémentation de l'architecture et de l'entraînement a été effectuée avec Tensorflow [89].

4.4. UNET-PROGRESSIVE

L'utilisation de convolutions dilatées est très bénéfique dans un CNN, puisque celles-ci permettent d'augmenter le champ récepteur sans engendrer de coût supplémentaire. UNet-Progressive est un autre CNN qui a été implémenté et qui utilise la dilatation. Initialement proposée pour effectuer la segmentation de tumeurs dans la vessie, cette architecture est basée sur le U-Net [65, 81]. Dans les prochaines sections, l'architecture du UNet-Progressive ainsi que des détails sur l'entraînement sont présentés.

4.4.1. Architecture

Tout comme l'architecture LungNet, UNet-Progressive est un CNN totalement convolutif qui produit une carte de segmentation par classe. Un schéma de l'architecture est présenté à la figure 4.7. Comme dans le réseau U-Net original, le réseau consiste en une partie encodeur suivie d'un décodeur. Dans le

UNet-Progressive, des convolutions dilatées sont utilisées dans le but d'augmenter le champ récepteur. Contrairement au réseau LungNet, les dilatations sont relativement petites, avec des facteurs de 1, 2 et 4 dans chacun des modules du côté encodeur. Ce choix de facteurs permet d'abord d'éviter les problèmes de maillage qui surviennent avec une augmentation trop agressive de la dilatation (voir la section 3.4.2.1). Aussi, augmenter la dilatation dans les 4 modules permet à ceux-ci d'extraire de l'information multiéchelles.

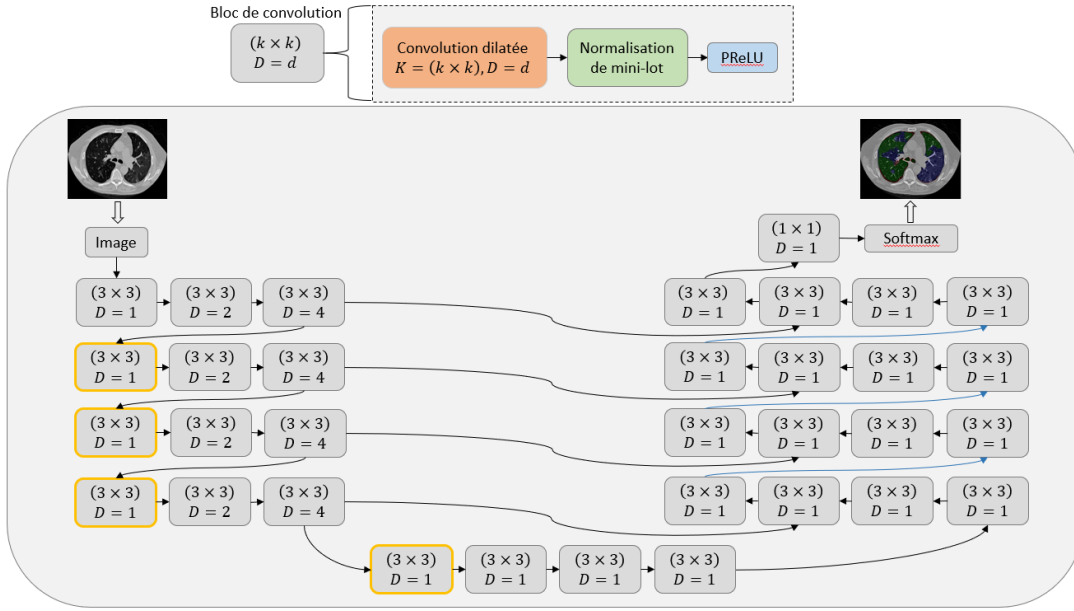


FIGURE 4.7. Le réseau UNet-Progressive est basé sur l'architecture U-Net. En haut, les composantes du bloc de convolution sont présentées, soit une couche de convolution ayant un noyau de dimensions K et un facteur de dilatation D , suivie d'une normalisation des cartes d'activations du mini-lot et d'une unité rectifiée linéaire paramétrique. Dans le réseau, du côté gauche, l'encodeur contient 4 modules (1 par ligne) et le facteur de dilatation D est augmenté progressivement à l'intérieur de chacun de ceux-ci de façon exponentielle. Chaque changement de module se fait avec un sous-échantillonnage d'un facteur 2 par l'augmentation du pas de convolution (blocs avec contours en orange). Du côté décodeur, chaque module contient 4 blocs basés sur des convolutions non dilatées et les changements de modules se font par l'application d'un sur-échantillonnage d'un facteur 2 (flèches bleues). Des connexions raccourcis sont également présentes entre les modules de mêmes niveaux.

L'autre mécanisme utilisé pour l'augmentation du champ récepteur est l'utilisation d'un pas de convolution de 2 lors du passage d'un module à un autre. Du côté encodeur et du pont (jonction entre l'encodeur et le décodeur), 16 blocs de convolutions sont implémentés, où chaque bloc consiste en une couche de convolution, suivie d'une normalisation de mini-lot et de l'application d'une fonction d'activation PReLU. Du côté décodeur, 17 blocs de convolution sont utilisés. Aussi, à la sortie d'un module, un sur-échantillonnage d'un facteur 2 est effectué pour retrouver progressivement les dimensions spatiales d'entrée. Comme dans l'architecture U-Net, des connexions raccourcis sont présentes entre les modules de niveaux pour intégrer de l'information de différents niveaux d'abstractions. À la fin du réseau, un bloc de convolution 1×1 est appliqué et une carte de sortie par classe est obtenue. Un dernier élément à noter est que, contrairement à LungNet, les dimensions spatiales et le nombre de cartes d'activations de chaque couche varient. Par exemple, du côté encodeur, chaque opération de sous-échantillonnage diminue les dimensions des cartes d'un facteur 2. Cela permet d'augmenter le nombre de cartes d'activations considérées et d'extraire davantage de caractéristiques, sans augmenter la complexité du réseau. Les différentes dimensions pour une image d'entrée de dimension 512×512 sont présentées dans le tableau 4.4.

4.4.2. Entraînement

L'entraînement du UNet-Progressive a été réalisé avec la même configuration de données que LungNet avec une validation croisée à 5 volets. Pour ce faire, l'entropie croisée, une fonction coût supervisée, a été minimisée avec l'optimiseur Adam. L'optimiseur a été utilisé avec un taux d'apprentissage initial de 10^{-4} et des valeurs standards de 0.9 et 0.99 pour les taux de décroissances β_1 et β_2 respectivement. Aussi, tel que proposé dans une étude similaire, lorsque le *DSC* calculé sur l'ensemble de validation ne s'améliore pas pendant 20 époques, le taux d'apprentissage est dégradé d'un facteur 2 [65]. Les poids initiaux sont générés par l'initialisation Glorot [90]. L'implémentation de l'architecture et de l'entraînement a été effectuée avec pyTorch [91].

TABLEAU 4.4. Dimensions spatiales pour une image d’entrée d’une dimension de 512×512 et nombre de cartes d’activations dans le UNet-Progressive pour les différents modules (Mod.) de l’encodeur (Enc.) et du décodeur (Déc.)

Section du réseau	Cartes	Dimensions
Module 1 (Enc.)	32	512×512
Module 2 (Enc.)	64	256×256
Module 3 (Enc.)	128	128×128
Module 4 (Enc.)	256	64×64
Pont	512	32×32
Mod. 1, blocs 1-2 (Déc.)	512	64×64
Mod. 1, bloc 3 (Déc.)	768	64×64
Mod. 1, bloc 4 (Déc.)	256	64×64
Mod. 2, blocs 1-2 (Déc.)	256	128×128
Mod. 2, bloc 3 (Déc.)	384	128×128
Mod. 2, bloc 4 (Déc.)	128	128×128
Mod. 3, blocs 1-2 (Déc.)	128	256×256
Mod. 3, bloc 3 (Déc.)	192	256×256
Mod. 3, bloc 4 (Déc.)	64	256×256
Mod. 4, blocs 1-2 (Déc.)	64	512×512
Mod. 4, bloc 3 (Déc.)	96	512×512
Mod. 4, bloc 4 (Déc.)	32	512×512
Sortie	3	512×512

4.4.2.1. *Entraînement avec régularisation*

Le nombre de paramètres à déterminer pendant l’entraînement dans le UNet-Progressive (2.0×10^7 paramètres) est considérablement plus grand que dans le réseau LungNet (1.3×10^5 paramètres). La quantité de données disponibles est relativement limitée, ce qui peut restreindre les performances du UNet-Progressive. Afin d’améliorer les performances de généralisation, plusieurs éléments de régularisation (voir la section 3.3) sont introduits dans le cadre d’une autre phase d’entraînement. D’abord, une augmentation de données est effectuée de sorte à augmenter artificiellement le nombre de données d’entraînement. Des opérations de rotation et de réflexion ont été appliquées aléatoirement, afin d’augmenter légèrement les données à chaque époque sans toutefois générer des images sur lesquelles les réalités de terrain ne seraient plus valides. Une autre technique de régularisation introduite est l’abandon bidimensionnel [92], utilisé afin de désactiver certaines cartes d’activations à la fin du pont durant l’entraînement, à chaque inférence.

Une fonction coût semi-supervisée et régularisée \mathcal{L} , donnée par l'équation 4.4.1, est minimisée avec l'optimiseur Adam.

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = -(1 - \chi) \frac{\sum_{j \in \mathcal{A}} \sum_{k=1}^{N_c} \gamma_{jk} y_{jk} \ln(\hat{y}_{jk})}{\sum_{j \in \mathcal{A}} \sum_{k=1}^{N_c} \gamma_{jk} y_{jk}} - \frac{\chi}{|\mathcal{N}|} \sum_{j \in \mathcal{N}} \sum_{k=1}^{N_c} \hat{y}_{jk} \ln(\hat{y}_{jk}) + \lambda \Omega(\mathbf{w}) \quad (4.4.1)$$

Dans la fonction coût, le premier terme est supervisé et correspond à l'entropie croisée catégorique pondérée entre les réalités de terrain et les activations softmax du réseau. Comme l'évaluation du terme requiert la connaissance des réalités de terrain, celui-ci est uniquement évalué pour l'ensemble des voxels annotés \mathcal{A} . Les poids contenus dans le vecteur γ permettent de pondérer le coût selon la classe de la prédiction, de sorte que les tissus sous-représentés (le poumon sain) contribuent au coût de façon comparable aux autres classes. Pour une classe donnée, le poids est inversement proportionnel à sa fréquence dans les données d'entraînement. Le deuxième terme est non supervisé et correspond à l'entropie moyenne des voxels non annotés \mathcal{N} . Des termes non supervisés basés sur l'entropie sont basés sur l'hypothèse des grappes (*clusters*) et ont été utilisés dans plusieurs applications pour améliorer les performances de généralisation [28, 93–95]. L'importance du terme non supervisé dans l'évaluation du coût est contrôlée par le facteur $\chi = 0.1$. Finalement, le troisième terme correspond à un terme de pénalité sur la norme des poids du réseau contrôlé par le facteur de régularisation $\lambda = 10^{-4}$. Dans cet entraînement, la régularisation L^2 a été utilisée, c'est-à-dire que $\Omega(\mathbf{w}) = \|\mathbf{w}\|_2^2$. Le réseau décrit dans cette section est référé en tant que UNet-Progressive-Reg.

Chapitre 5

RÉSULTATS

Au cours de cette section, les différents résultats obtenus à la suite de la méthodologie du chapitre 4 sont présentés. Les résultats sont regroupés en trois sections distinctes. D’abord, la segmentation automatique est évaluée par une vérification visuelle des cartes de sortie et des segmentations. Ensuite, les performances des différents CNN sont évaluées sur des données partiellement annotées, similaires aux données qui sont utilisées durant l’entraînement. Les annotations partielles, qui proviennent de la base de données du HUG, permettent d’évaluer la validité des prédictions à l’intérieur de celles-ci. Comme ces annotations sont partielles, il est possible que d’autres tissus de la même classe se trouvent à l’extérieur de l’annotation de référence. Finalement, l’utilisation des CNN dans un contexte plus près des applications cliniques envisagées est évaluée. Pour ce faire, les patients complètement annotés sont introduits afin d’intégrer plusieurs types de tissus normalement exclus des annotations partielles, mais qui seraient rencontrés dans une application réelle.

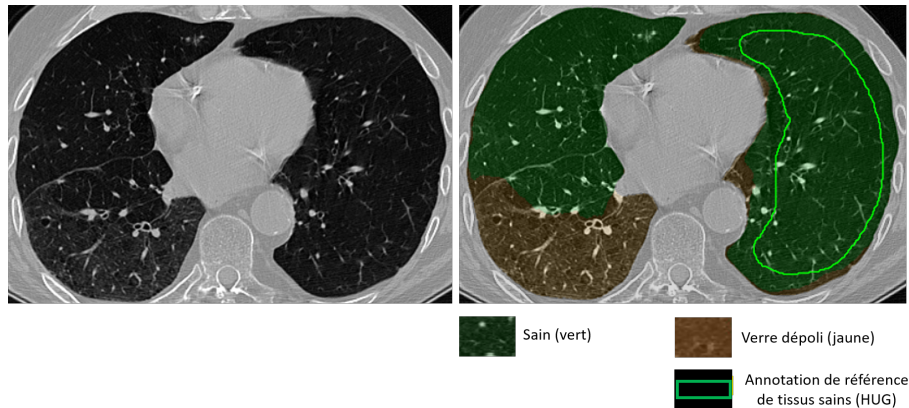


FIGURE 5.1. Une segmentation pour laquelle une annotation partielle saine est disponible, effectuée par le UNet-Progressive-Reg. L’image de gauche est traitée afin de donner en sortie la carte de segmentation de droite. La carte de couleurs superposée aux poumons, en arrière-plan, correspond aux prédictions du réseau pour chaque voxel (poumon sain en vert, verre dépoli en jaune et fibrose en rouge). Le contour en vert correspond à la région où l’annotation partielle de référence saine est disponible.

5.1. CARTES DE SEGMENTATION

Dans cette section, des cartes de segmentation pouvant être obtenues avec les différents CNN sont présentées. Particulièrement dans un contexte où les annotations ne sont pas complètes, une représentation visuelle des segmentations fournit une information complémentaire aux métriques qui est essentielle pour évaluer les prédictions du réseau. En effet, cela permet d’apprécier la totalité des prédictions faites sur une coupe, ce qui n’est pas le cas avec les métriques calculées sur les annotations partielles. Dans le même ordre d’idées, il est possible de prendre conscience de la façon dont les CNN segmentent certaines régions et motifs en particulier. Dans un premier temps, la figure 5.1 présente la segmentation d’une coupe ayant reçu des annotations de tissus sains sur la majorité du poumon gauche. La segmentation est effectuée par le UNet-Progressive-Reg pour 3 classes (Sain, Verre dépoli et Fibrose). Comme les annotations sont partielles, certains tissus hors de l’annotation saine peuvent également être sains. Une inspection visuelle permet de constater que la majorité du poumon gauche semble également saine. Dans le même ordre d’idées, l’apparence des motifs présents en bas de ce poumon correspond à du verre dépoli.

Une segmentation pour laquelle une petite réalité de terrain de verre dépoli est disponible est présentée à la figure 5.2. La coupe segmentée contient des motifs de verre dépoli et de fibrose diversifiés et relativement complexes. Il y a également une présence importante de l’arbre trachéobronchique. Il est important de noter que dans certains cas, les régions annotées peuvent être très petites comparativement aux dimensions de la région pulmonaire.

Deux coupes où il y a une présence de fibrose sont segmentées aux figures 5.3 et 5.4. La coupe présentée à la figure 5.3 est sévèrement atteinte de fibrose, tandis que les motifs illustrés à la figure 5.4 sont plus subtils. Les segmentations d’une coupe, effectuées par les CNN UNet-Progressive-Reg, UNet-Progressive et LungNet sont données à la figure 5.5. Il est possible d’y voir que malgré une assignation presque parfaite des voxels à l’intérieur des annotations partielles, les segmentations diffèrent à l’extérieur de celles-ci.

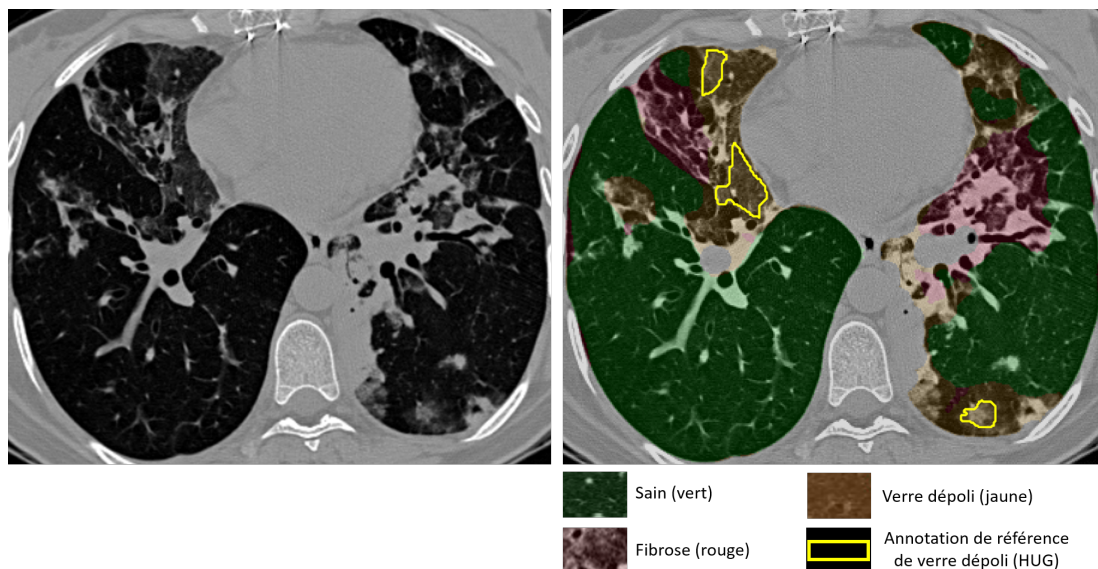


FIGURE 5.2. Une segmentation pour laquelle une réalité de terrain de verre dépoli est disponible, effectuée par le UNet-Progressive-Reg. Les zones annotées sont petites et la majorité des voxels pulmonaires n'ont pas reçu de réalité de terrain. Le contour en jaune correspond à la région pour laquelle une annotation partielle de référence de verre dépoli est disponible.

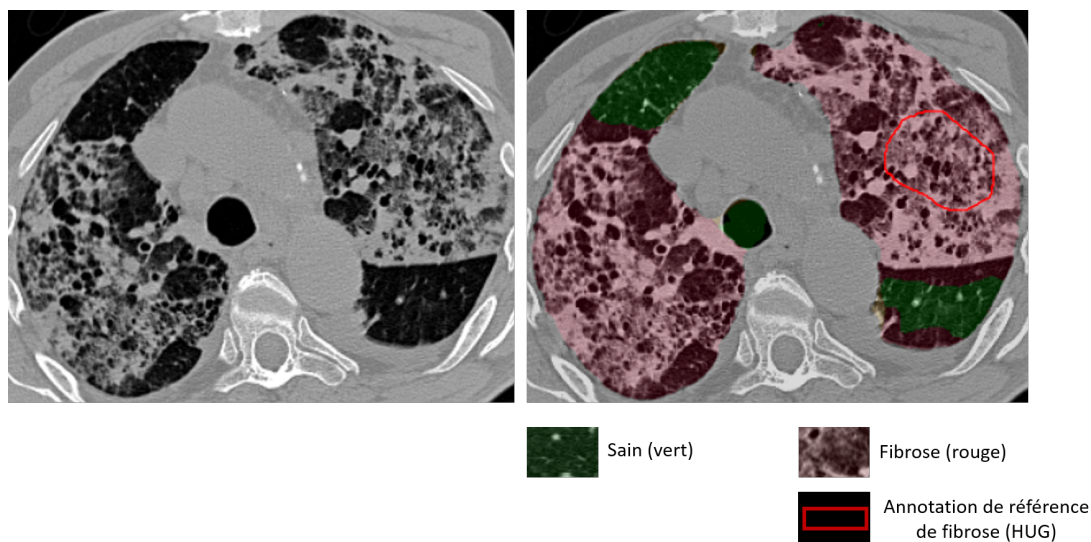


FIGURE 5.3. Une segmentation pour laquelle une réalité de terrain de fibrose est disponible, effectuée par le UNet-Progressive-Reg. De la fibrose sévère couvre la majorité des poumons. Les poumons contiennent essentiellement 2 types de tissus distincts (poumon sain et fibrose sévère). Le contour en rouge correspond à la région pour laquelle une annotation partielle de référence de fibrose est disponible.

Les fluctuations entre les cartes produites sont légères et touchent essentiellement l'étendue du verre dépoli dans le poumon gauche. Une segmentation où les cartes diffèrent davantage est donnée à la figure 5.6. Alors que la majorité des voxels annotés sont bien assignés, il est possible de voir une différence importante au niveau de l'étendue du poumon sain.

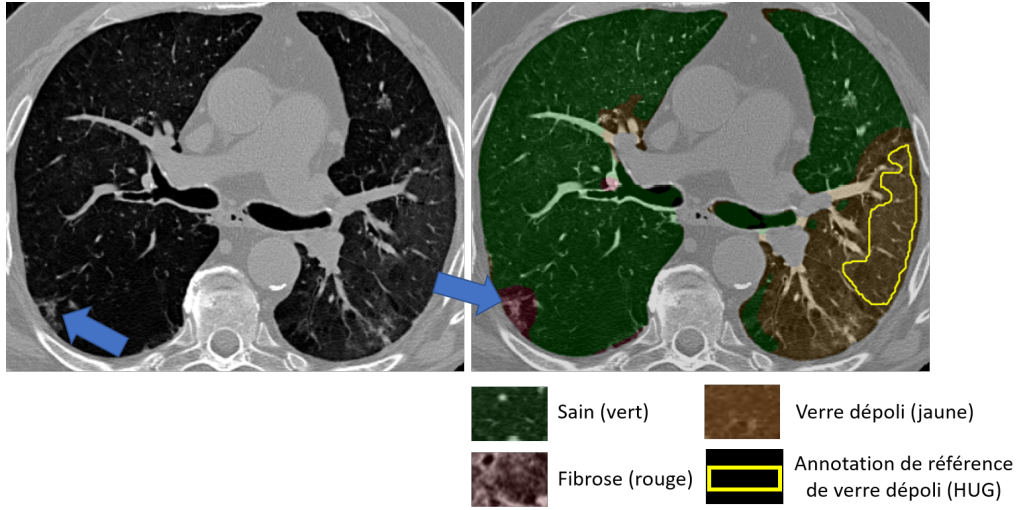


FIGURE 5.4. Une segmentation pour laquelle une annotation partielle de référence de verre dépoli est disponible (contour jaune) est présentée. En bas à gauche, un motif de fibrose plus subtile est détecté par le CNN (flèches bleues).

À la suite du passage d'une image dans le réseau, un volume softmax de dimensions $512 \times 512 \times 3$ est généré. Les différentes cartes représentent la probabilité d'appartenir aux classes Sain, Verre dépoli et Fibrose pour chacun des voxels. Les cartes softmax des trois coupes segmentées aux figures 5.1 à 5.3 sont données aux figures 5.7, 5.8 et 5.9. Aussi, la carte softmax des prédictions, c'est-à-dire la carte des activations liées aux classes assignées, est présentée pour ces coupes. Cette carte permet de détecter les régions où les softmax sont plus faibles, indiquant une confusion entre au moins 2 classes.

Les figures 5.7 et 5.8 sont des exemples où la majorité des activations softmax des prédictions sont élevées. Les régions caractérisées par des activations softmax plus faibles correspondent généralement aux interfaces entre les régions de classes différentes. Les cartes de la figure 5.9 sont le résultat du passage d'une coupe relativement plus complexe dans le réseau. Les activations softmax

des prédictions demeurent élevées dans la plupart des régions, mais une confusion entre les tissus de verre dépoli et de fibrose survient dans certaines régions, notamment près de l'arbre trachéobronchique.

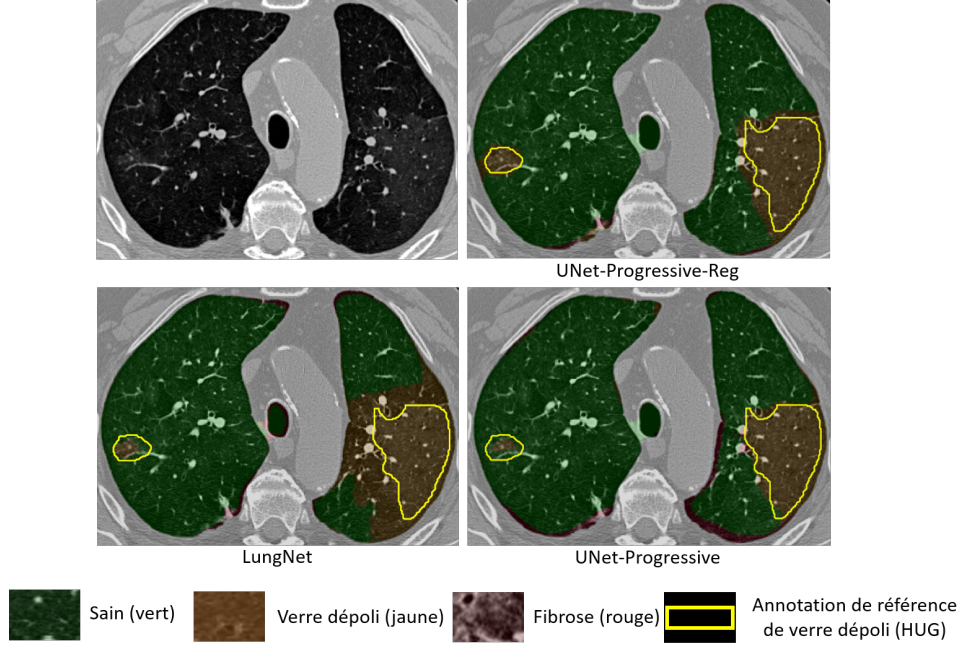


FIGURE 5.5. Une coupe (en haut à gauche) est segmentée par les CNN UNet-Progressive-Reg (en haut à droite), LungNet (en bas à gauche) et UNet-Progressive (en bas à droite). Les contours en jaune correspondent aux annotations partielles de verre dépoli. Hors de ces contours, de légères différences au niveau de l'étendue de verre dépoli (poumon droit) sont perceptibles.

5.2. PERFORMANCES DE LA SEGMENTATION AUTOMATIQUE

Dans cette section, les performances de l'algorithme de segmentation sont évaluées. Pour ce faire, les matrices de confusion et les métriques discutées au cours de la section 4.2.2 sont présentées pour les données de validation. Ces métriques permettent d'évaluer les segmentations sur des données similaires à celles fournies au réseau durant l'entraînement, soit des coupes avec des réalités de terrain qui contiennent une grande diversité de motifs et de types de tissus.

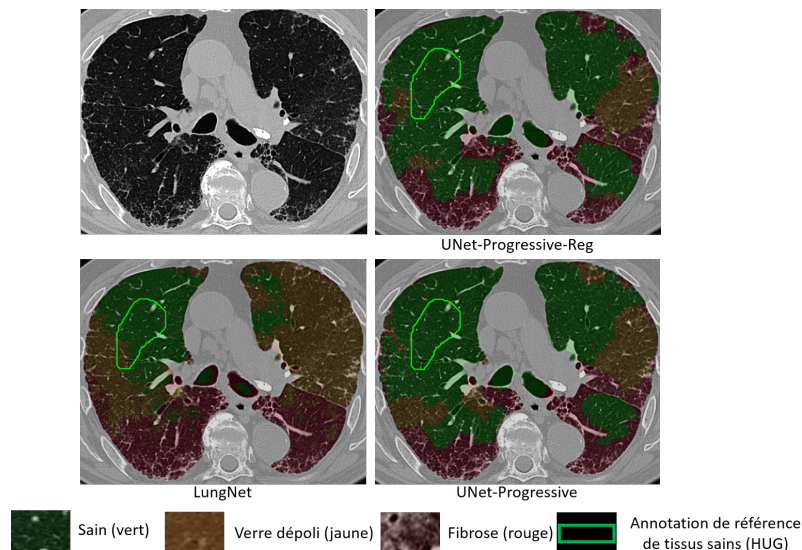


FIGURE 5.6. Une coupe (en haut à gauche) est segmentée par les CNN UNet-Progressive-Reg (en haut à droite), LungNet (en bas à gauche) et UNet-Progressive (en bas à droite). Les contours en vert correspondent aux annotations partielles de référence de poumon sain. Hors de ces contours, des différences importantes au niveau de l'étendue de poumon sain (sur les deux poumons) sont visibles.

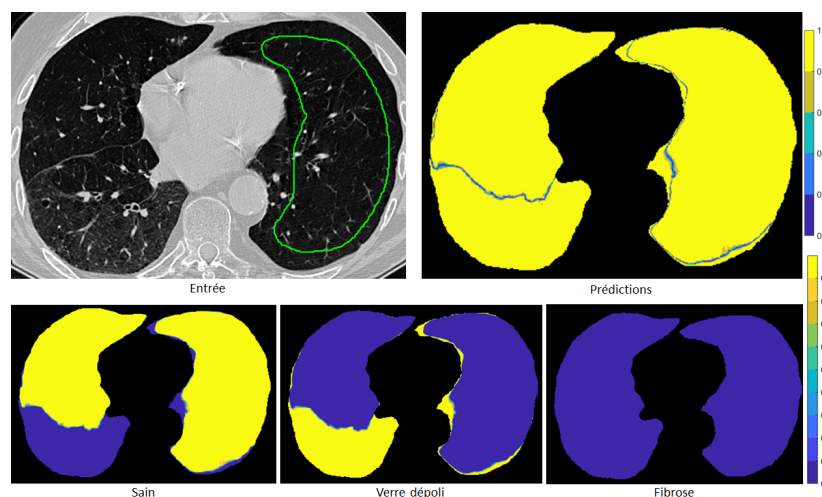


FIGURE 5.7. Cartes d'activations softmax à la suite du passage d'une image avec une annotation de référence partielle (contour vert) de poumon sain du HUG (en haut à gauche) dans le UNet-Progressive-Reg. En bas, les cartes d'activations softmax des classes Sain, Verre dépoli et Fibrose sont présentées à gauche, au centre et à droite respectivement. La carte des activations softmax des prédictions est également présentée en haut à droite.

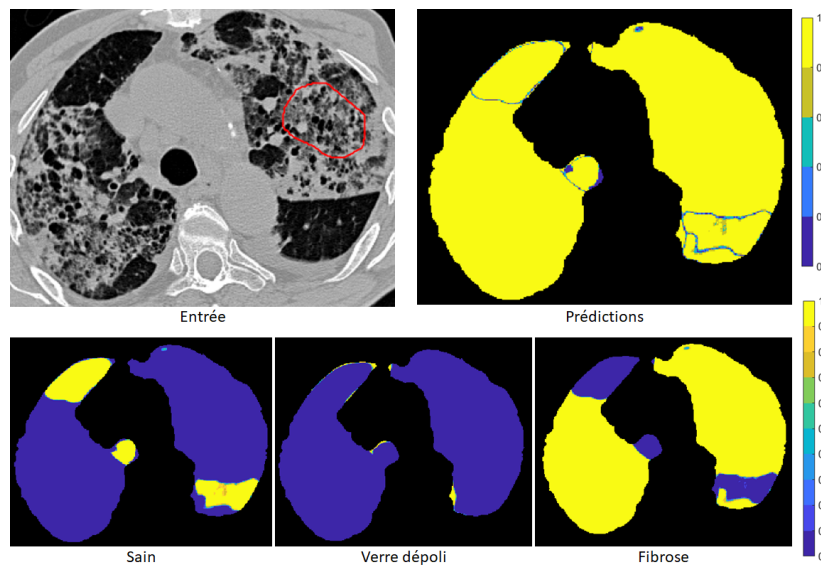


FIGURE 5.8. Cartes d'activations softmax à la suite du passage d'une image avec annotation de référence partielle de fibrose du HUG (en haut à gauche) dans le UNet-Progressive-Reg. Les cartes d'activations softmax des classes Sain, Verre dépoli et Fibrose ainsi que des activations softmax des prédictions sont présentées.

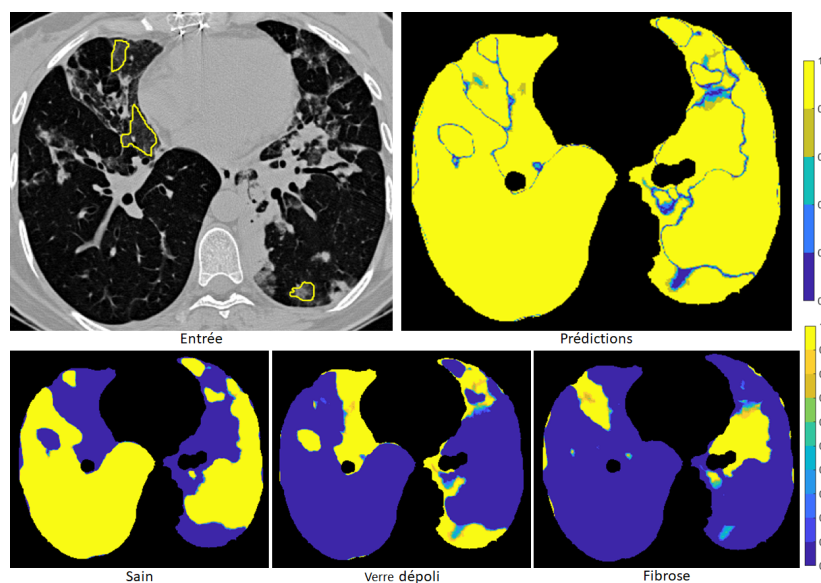


FIGURE 5.9. Cartes d'activations softmax suite au passage d'une image contenant des annotations partielles de référence (contours jaunes) de verre dépoli et des motifs diversifiés et subtils (en haut à gauche) dans le UNet-Progressive-Reg. Les cartes d'activations softmax des classes Sain, Verre dépoli et Fibrose ainsi que des activations softmax des prédictions sont illustrées.

Dans un premier temps, des matrices de confusion pour différents volets sont présentées à la figure 5.10. Ces matrices de confusion permettent d'évaluer la fraction des voxels de chacune des classes qui ont été attribués aux différents types d'assignations. Celles-ci ont été évaluées à l'intérieur des annotations de référence, étant donné que les délimitations du HUG ne sont pas complètes. Par exemple, la valeur (1,3) de ces matrices s'interprète comme la fraction des voxels sains qui ont été assignés comme étant de la fibrose. Les matrices sont données pour les trois CNN discutés dans la méthodologie (LungNet, UNet-Progressive, UNet-Progressive-Reg). La justesse d'assignation A correspond à la fraction des voxels qui a été assignée correctement par rapport aux annotations des radiologues. Le coefficient Dice s'interprète comme l'évaluation de la similarité entre une segmentation prédite et la réalité de terrain. Cette métrique est pénalisée par les faux positifs et les faux négatifs de façon équivalente. Les coefficients Dice pour le poumon sain (DSC_S), le verre dépoli (DSC_{VD}) et la fibrose (DSC_F) ainsi que la justesse d'assignation ont été calculés. L'évaluation de ces deux métriques pour les volets présentés à la figure 5.10 est donnée dans le tableau 5.1.

TABLEAU 5.1. Coefficients Dice (DSC), précisions (Pr) et recouvrements (Re) pour les 3 classes Fibrose (F), Verre dépoli (VD) et Sain (S), ainsi que les justesses d'assignation (A) pour différents volets V des CNN LungNet (LN), UNet-Progressive (UP) et UNet-Progressive-Reg (UPR)

V	CNN	A	DSC_F	DSC_{VD}	DSC_S	Pr_F	Pr_{VD}	Pr_S	Re_F	Re_{VD}	Re_S
(1)	LN	0.96	0.88	0.95	0.98	0.85	0.95	0.99	0.90	0.95	0.98
	UP	0.98	0.92	0.97	0.99	0.91	0.98	0.99	0.93	0.96	0.99
	UPR	0.98	0.94	0.98	0.99	0.92	0.98	0.99	0.96	0.98	0.99
(2)	LN	0.88	0.86	0.86	0.94	0.97	0.77	0.99	0.78	0.97	0.89
	UP	0.87	0.86	0.85	0.95	0.93	0.81	0.91	0.80	0.89	0.99
	UPR	0.92	0.92	0.89	0.95	0.93	0.89	0.93	0.91	0.89	0.98
(3)	LN	0.82	0.83	0.69	0.98	0.84	0.69	0.96	0.84	0.69	1
	UP	0.78	0.81	0.64	0.94	0.81	0.64	0.92	0.80	0.64	0.97
	UPR	0.88	0.88	0.81	0.99	0.93	0.75	1	0.83	0.90	0.98

Les précisions et les recouvrements ont aussi été évalués pour le poumon sain (Pr_S et Re_S), le verre dépoli (Pr_{VD} et Re_{VD}) et la fibrose (Pr_F et Re_F). La précision d'un modèle est pénalisée par les faux positifs qu'il émet tandis que le recouvrement est pénalisé par les faux négatifs. Les précisions et recouvrements

des trois CNN sont également donnés dans le tableau 5.1. De plus, les différentes métriques ont été calculées pour l'ensemble de la validation croisée. Ces métriques sont utilisées afin d'apprécier les performances sur un grand nombre de patients.

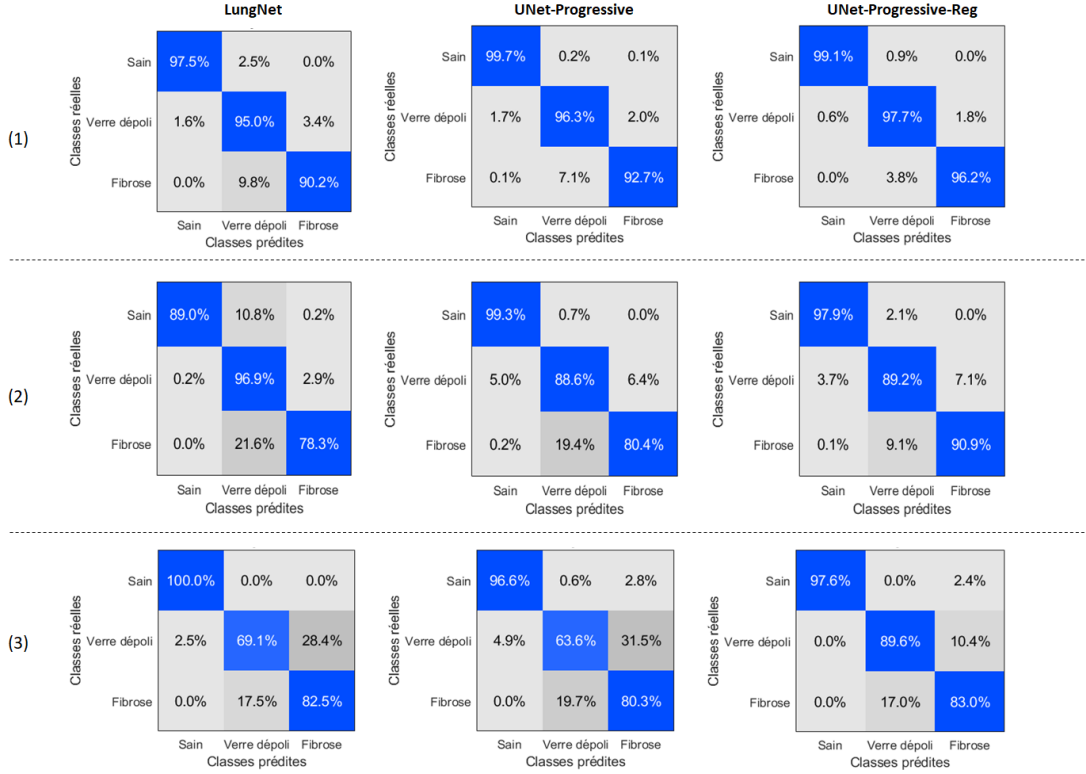


FIGURE 5.10. Les matrices de confusion 3 classes des LungNet, UNet-Progressive et UNet-Progressive-Reg pour 3 volets (dénnotés (1), (2) et (3)). Les différentes valeurs de la matrice représentent la fraction des voxels ayant la classe réelle associée à la ligne. Par exemple, la valeur (1,2) de la matrice de confusion LungNet (1) indique que 2.5 % des voxels ont été assignés comme étant du verre dépoli, alors que les annotations pour ces voxels sont du poumon sain.

Dans un premier temps, la matrice de confusion de chacun des réseaux est présentée à la figure 5.11. Aussi, les métriques moyennes dérivées des matrices de confusion ainsi que leurs écarts-types sont incluses dans le tableau 5.2. Le tableau 5.3 présente les activations softmax moyennes des voxels correctement et incorrectement assignés, ainsi que pour toutes les prédictions. Cela est fait dans l'optique d'évaluer la faisabilité d'ajouter un seuil softmax pour réduire davantage les assignations erronées. La figure 5.12 illustre un exemple de situation qui pourrait bénéficier d'un tel seuil.

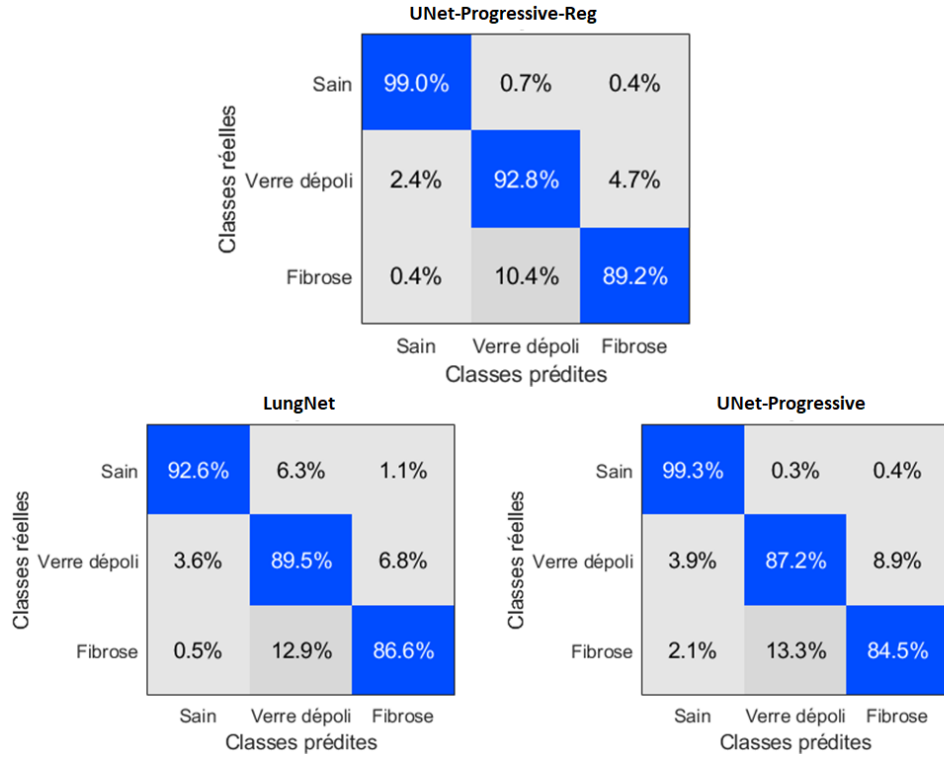


FIGURE 5.11. Les matrices de confusion 3 classes de LungNet, UNet-Progressive et UNet-Progressive-Reg pour toutes les données de validation. Les différentes valeurs contenues dans la matrice correspondent à la fraction des voxels ayant la classe réelle associée à la ligne.

TABLEAU 5.2. Coefficients Dice (DSC), précisions (Pr), recouvrements (Re) et justesses d'assignation (A) des UNet-Progressive, LungNet et UNet-Progressive-Reg sur les données de validation

Métrique	UNet-Progressive	LungNet	UNet-Progressive-Reg
A	0.88 ± 0.07	0.89 ± 0.05	0.93 ± 0.04
Re_F	0.85 ± 0.05	0.87 ± 0.07	0.89 ± 0.06
Re_{VD}	0.84 ± 0.12	0.87 ± 0.11	0.92 ± 0.03
Re_S	0.99 ± 0.01	0.94 ± 0.06	0.99 ± 0.01
Pr_F	0.90 ± 0.05	0.90 ± 0.06	0.94 ± 0.02
Pr_{VD}	0.83 ± 0.13	0.81 ± 0.10	0.87 ± 0.09
Pr_S	0.94 ± 0.03	0.96 ± 0.05	0.97 ± 0.03
DSC_F	0.87 ± 0.04	0.88 ± 0.04	0.91 ± 0.03
DSC_{VD}	0.84 ± 0.12	0.84 ± 0.09	0.90 ± 0.06
DSC_S	0.96 ± 0.02	0.95 ± 0.05	0.98 ± 0.02

TABLEAU 5.3. Activations softmax moyennes et écarts-types σ_{So} de tous les volets selon les assignations effectuées (valides, erronées, toutes les assignations)

Assignations	So	σ_{So}
Valide	0.97	0.08
Erronée	0.86	0.19
Toutes	0.96	0.10

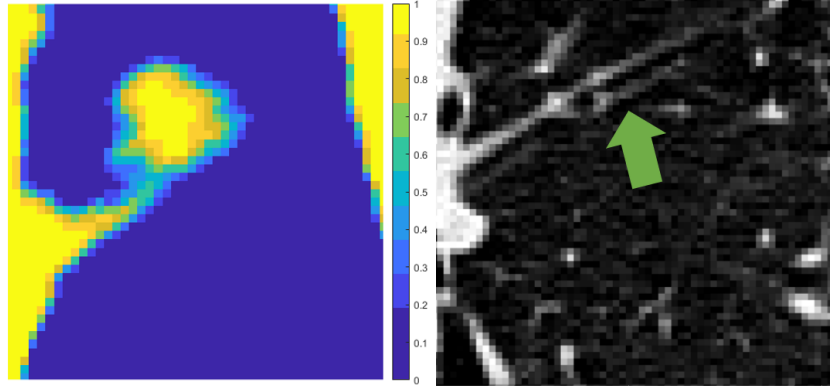


FIGURE 5.12. La région d'intérêt (à droite) contient du tissu sain. Toutefois, une partie de cette région contient des activations softmax de fibrose (à gauche) d'amplitudes moyennes (teintes de cyan), suffisamment élevées pour que les voxels soient assignés incorrectement en tant que fibrose. Cette région est un exemple de situation qui pourrait bénéficier d'un seuil softmax, où aucune assignation ne serait effectuée pour les voxels ayant une activation trop faible.

5.3. ÉVALUATION DE FUTURES APPLICATIONS

Cette section présente des résultats en lien avec l'évaluation de futures applications en radio-oncologie. Tel que mentionné précédemment, une application envisageable est de faciliter l'évaluation de la fibrose par les radio-oncologues en effectuant préalablement une segmentation complète des coupes avec un CNN.

Dans cette optique, les 3 patients qui ont été complètement annotés par les radio-oncologues du CHUM sont utilisés afin d'évaluer le UNet-Progressive-Reg dans un contexte plus près des applications cliniques. Les annotations manuelles ont été effectuées pour 2 classes, soit les tissus sains et les tissus pouvant être associés à de la fibrose pulmonaire.



Image test

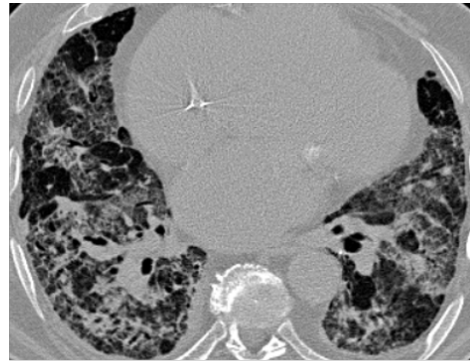
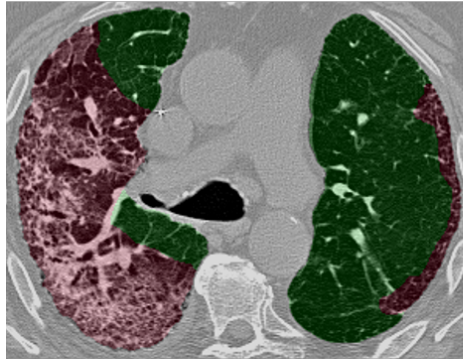
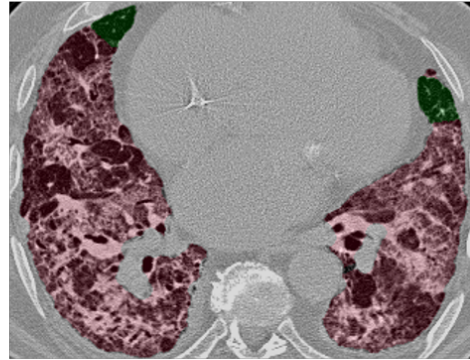


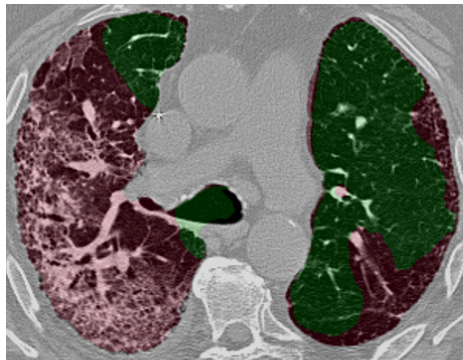
Image test



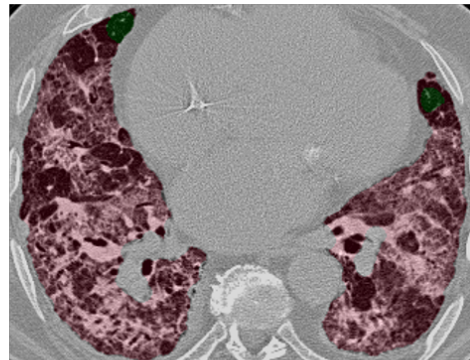
Vérité terrain



Vérité terrain



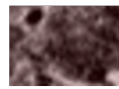
Prédiction



Prédiction



Sain (vert)



Fibrose (rouge)

FIGURE 5.13. Deux images du patient 1 (en haut) et leurs segmentations (en bas) sont présentées. Au centre, les annotations complètes de référence du CHUM sont illustrées. Les voxels en rouge correspondent aux tissus pouvant être associés à de la fibrose pulmonaire et les voxels en vert représentent les tissus sains. Une fraction importante du poumon gauche est saine sur la majorité des coupes.

Les 3 patients qui ont été complètement annotés sont atteints par la fibrose pulmonaire à différents degrés. Deux segmentations du patient 1 sont présentées à la figure 5.13 avec les réalités de terrain complètes associées.

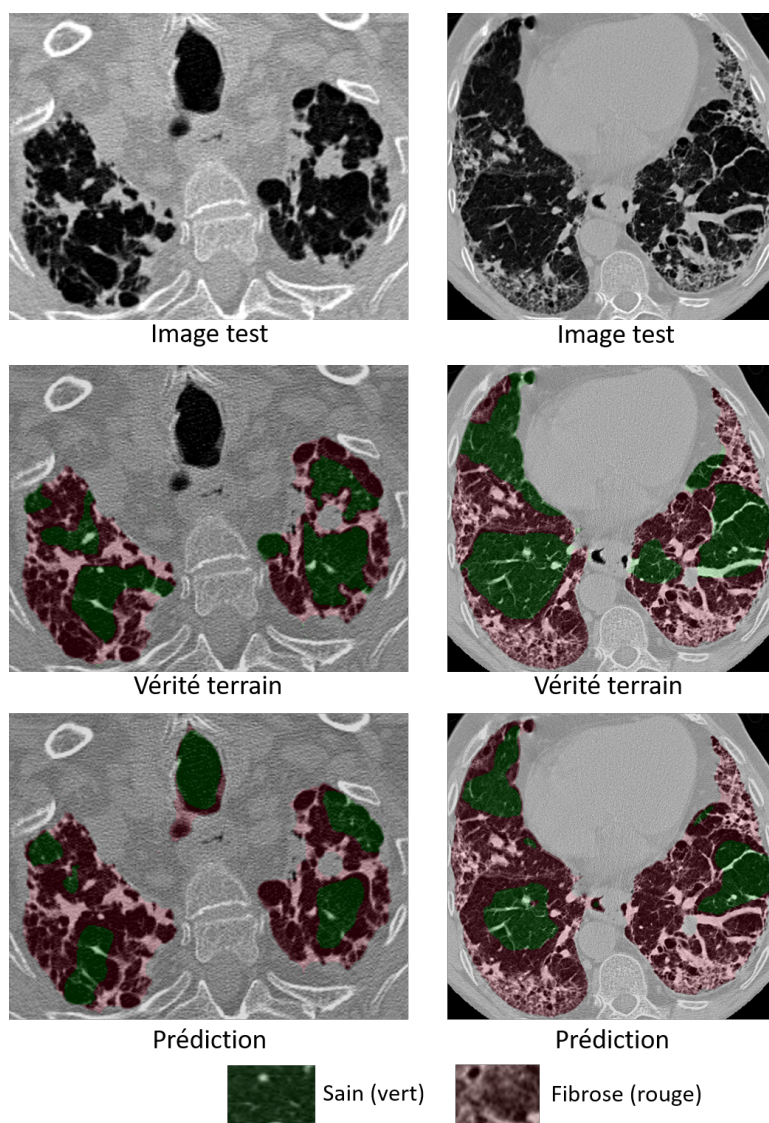


FIGURE 5.14. Deux images du patient 2 (en haut) et leurs segmentations (les tissus de fibrose pulmonaire en rouge et le poumon sain en vert) sont présentées (en bas). Au centre, les annotations complètes de référence du CHUM sont illustrées. Le patient 2 contient également plusieurs régions saines. Il contient aussi d'autres pathologies, tels que des motifs d'emphysème.

Le patient 1 contient plusieurs coupes où le poumon droit est majoritairement sain, ainsi que différents degrés de sévérité de fibrose. Celui-ci représente un scénario où l'atteinte de fibrose est moyenne.

TABLEAU 5.4. Coefficients Dice (DSC), précisions (Pr), recouvrements (Re) et justesses d'assignation (A) du UNet-Progressive-Reg sur les 3 patients de test complètement annotés

Métrique	Tissu	Patient 1	Patient 2	Patient 3
DSC	Sain	0.77	0.83	0.39
	RFP	0.89	0.81	0.97
Re	Sain	0.64	0.75	0.25
	RFP	0.99	0.94	0.99
Pr	Sain	0.99	0.94	0.91
	RFP	0.81	0.73	0.95
A		0.85	0.83	0.95

La figure 5.14 illustre deux segmentations du patient 2, ainsi que les réalités de terrain complètes associées. Les motifs présents sur les coupes du patient 2 sont diversifiés, avec notamment une présence d'emphysème. Ce patient représente un cas où plusieurs motifs d'ILD sont présents. Finalement, des segmentations du patient 3 sont données à la figure 5.15. Ce patient est caractérisé par une présence importante de fibrose sévère et très peu de tissus sains. Les régions identifiées comme étant saines sont très petites et entourées de régions de fibrose sévère. Le patient 3 correspond à un cas où la grande majorité des poumons est atteinte de fibrose.

TABLEAU 5.5. Coefficients Dice (DSC), précisions (Pr) et recouvrements (Re) du UNet-Progressive-Reg sur l'ensemble de test

Métrique	Sain	RFP
DSC	0.67 ± 0.24	0.89 ± 0.08
Re	0.54 ± 0.26	0.98 ± 0.03
Pr	0.95 ± 0.04	0.83 ± 0.11

À partir des réalités de terrain complètes qui ont été produites manuellement, les performances ont pu être évaluées sur l'ensemble des poumons des coupes

disponibles. Les coefficients Dice, les recouvrements, les précisions ainsi que la justesse d'assignation sont donnés dans le tableau 5.4 pour chacun des 3 patients de test. Les métriques évaluées sur l'ensemble de test, pour lequel une justesse d'assignation de $A = 0.88 \pm 0.06$ a été obtenue, sont données dans le tableau 5.5.

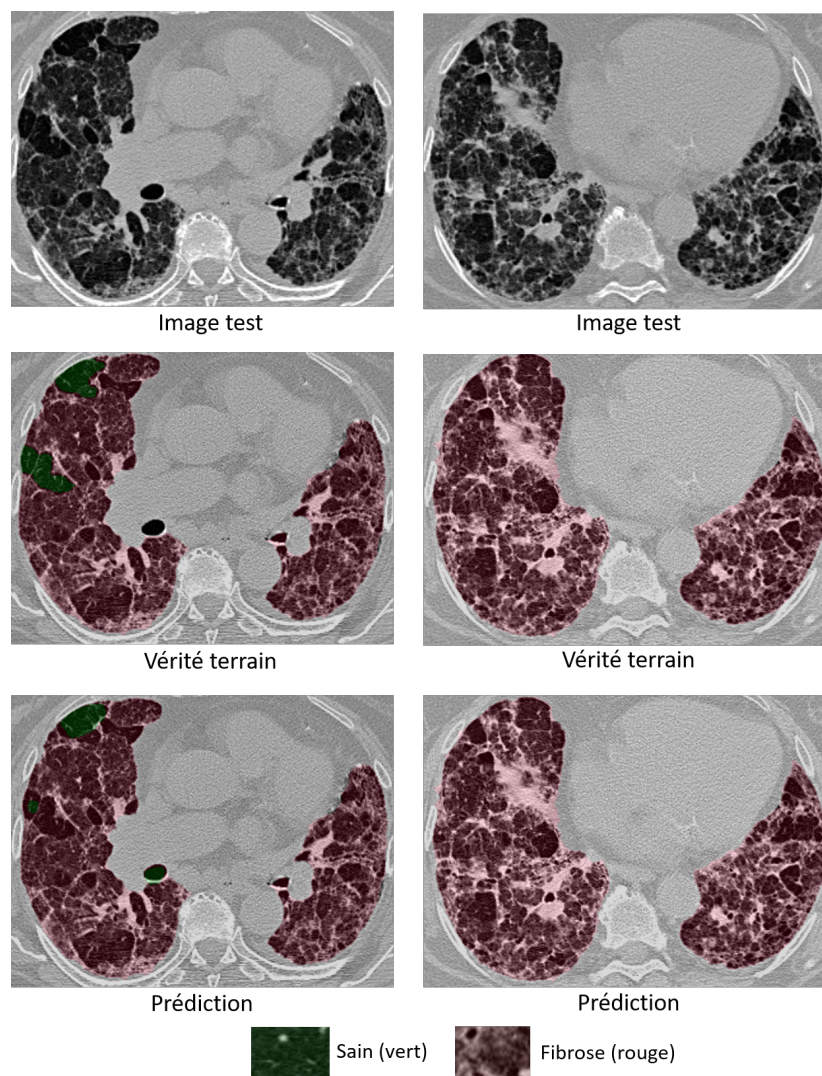


FIGURE 5.15. Deux images du patient 3 (en haut) et leurs segmentations (les tissus de fibrose pulmonaire en rouge et le poumon sain en vert) sont présentées (en bas). Au centre, les annotations complètes de référence du CHUM sont illustrées. L'étendue de la fibrose couvre la grande majorité des poumons et les régions annotées comme étant saines sont petites et entourées de fibrose sévère.

Chapitre 6

DISCUSSION

Dans ce chapitre, les différents résultats présentés précédemment sont discutés. Aussi, à la lumière des résultats obtenus, diverses applications qui pourraient être développées en radio-oncologie sont abordées.

D’abord, comme il est possible de le voir sur les exemples des figures 5.1, 5.2 et 5.3, les CNN peuvent produire des cartes de segmentation cohérentes. Par inspection visuelle, plusieurs constatations initiales peuvent être faites. En considérant la figure 5.1, qui représente un cas majoritairement sain avec présence de verre dépoli, il est d’abord possible d’observer que le poumon sain est assigné correctement à l’intérieur de la réalité de terrain partielle. Le réseau entraîné distingue bien les vaisseaux, qui pourraient être confondus avec de la fibrose étant donné leur densité élevée (figure 6.1).

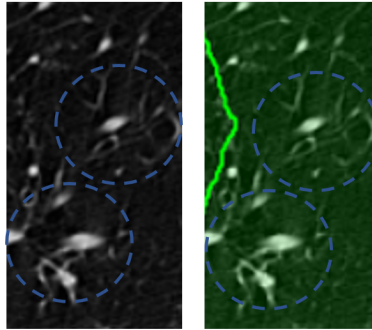


FIGURE 6.1. Sur la région d’intérêt extraite de la coupe présentée à la figure 5.1 (à gauche), plusieurs zones sont susceptibles d’être assignées comme de la fibrose étant donné la présence de plusieurs vaisseaux (cercles bleus). Le réseau entraîné assigne adéquatement ces motifs comme du poumon sain (à droite).

De plus, le réseau distingue correctement la présence de verre dépoli dans le bas du poumon droit. La figure 6.2 présente la frontière établie par le réseau entre les tissus sains et le verre dépoli. Cette interface se situe dans la région où a lieu un changement de densité perceptible sur l’image qui caractérise le verre dépoli.

De façon similaire, le modèle peut faire une distinction entre le verre dépoli et des motifs de fibrose plus subtils tels que la réticulation. Par exemple, à la figure 6.3, une région tirée de la figure 5.2 présente des prédictions plus complexes effectuées par le réseau. Par contre, bien que le modèle distingue bien la fibrose sévère et les tissus sains, les frontières entre ces 2 types de tissus ont tendance à inclure une bordure de tissus sains assignés comme de la fibrose. Un exemple de cette situation est présenté à la figure 6.4.

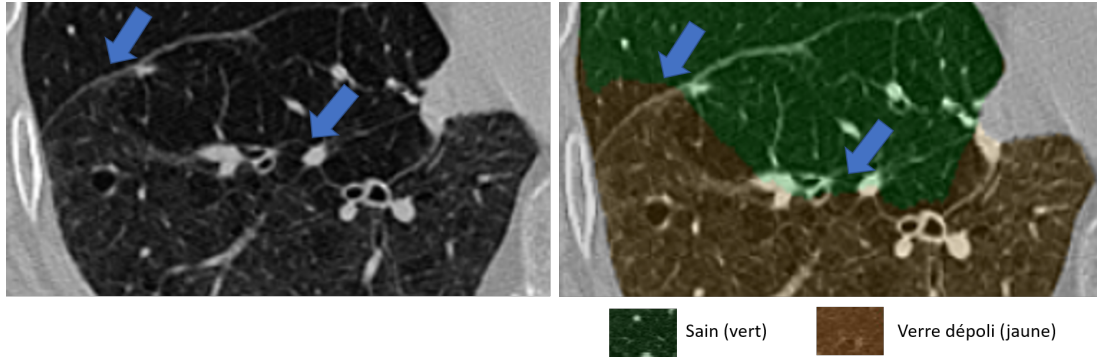


FIGURE 6.2. Sur la région d'intérêt extraite de la coupe présentée à la figure 5.1 (à gauche), il est possible de voir une zone où se fait un changement brusque de densité perceptible sur l'image, qui caractérise une transition entre des tissus sains et du verre dépoli (flèches bleues). La prédiction du réseau (à droite) établit une frontière entre les deux types de tissus (tissus sains en vert et verre dépoli en jaune) qui suit cette transition.

Une autre région où des assignations erronées surviennent est près de l'arbre trachéobronchique (figure 6.5). En effet, alors que les motifs présents ne correspondent pas à une présence d'ILD, ceux-ci ont tendance à être assignés comme étant de la fibrose. Ceci peut être dû au manque de données d'entraînement qui contiennent des annotations autour de l'arbre trachéobronchique ainsi qu'aux similitudes avec certains motifs de fibrose.

Pour un voxel donné, la classe assignée par un CNN correspond à celle associée à l'activation softmax la plus élevée. Dans le cas présent, les valeurs moyennes de softmax de prédictions valides sont très élevées ($So = 0.97 \pm 0.08$). Cela

signifie que lorsque le réseau effectue une prédiction et que celle-ci est exacte par rapport à l'annotation d'expert, alors la classe assignée se distingue nettement des autres classes considérées.

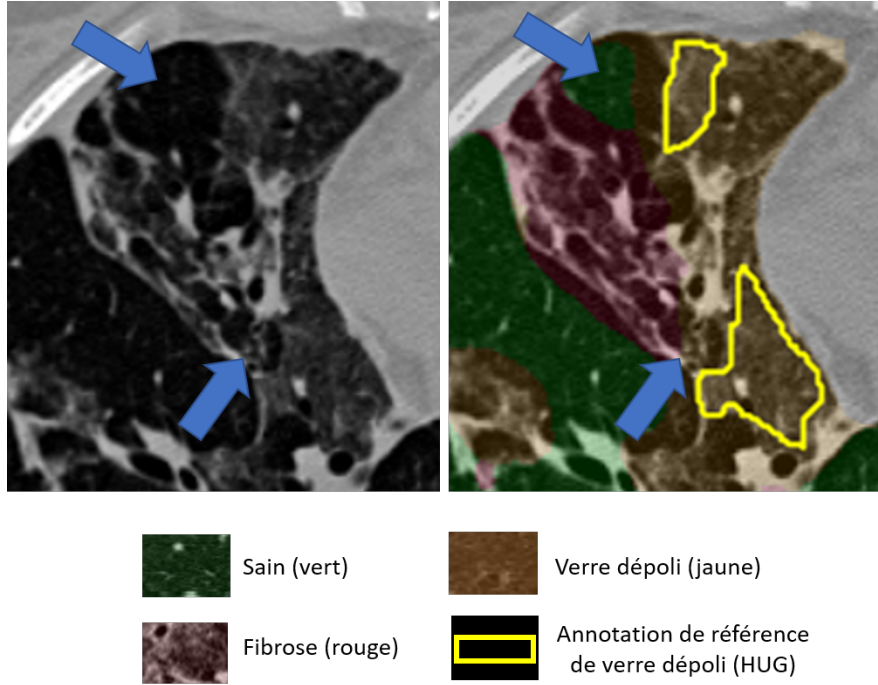


FIGURE 6.3. Sur la région d'intérêt extraite de la coupe présentée à la figure 5.2 (à gauche), il est possible de voir des exemples de motifs plus subtils. La flèche du haut montre une très petite région saine correctement assignée, susceptible de ne pas être détectée étant donné la présence environnante de verre dépoli et de fibrose. La flèche du bas indique une interface subtile entre du verre dépoli intense et de la fibrose établie par le réseau.

Cette constatation est importante puisque cela indique que lorsque les prédictions sont justes, il y a très peu de confusion entre les différentes classes. Dans les cas où le réseau émet des prédictions erronées, les activations softmax sont plus petites ($So = 0.86 \pm 0.19$), ce qui indique que le modèle est moins confiant lorsqu'il génère des erreurs dans les segmentations. De plus, les activations softmax fluctuent de façon considérable, ce qui suggère que les assignations erronées peuvent être le résultat de deux activations softmax similaires pour un même voxel et d'une confusion entre deux types de tissus. La figure 6.6 illustre un exemple où une confusion entre deux classes est présente.

Étant donné ces considérations sur les activations softmax, il pourrait être intéressant de considérer l'ajout d'un seuil minimal pour qu'une prédiction du réseau soit utilisée. Cela pourrait potentiellement réduire davantage le nombre d'assignations erronées. Pour ce faire, les voxels qui auraient reçu des activations softmax de prédictions trop petites ne recevraient simplement pas d'assignation sur les cartes de segmentation finales. Un exemple de situation pouvant bénéficier d'un tel seuil est présenté à la figure 5.12. Toutefois, tel que suggéré par le softmax moyen des prédictions ($So = 0.96 \pm 0.10$), ces situations ne sont pas couramment rencontrées.

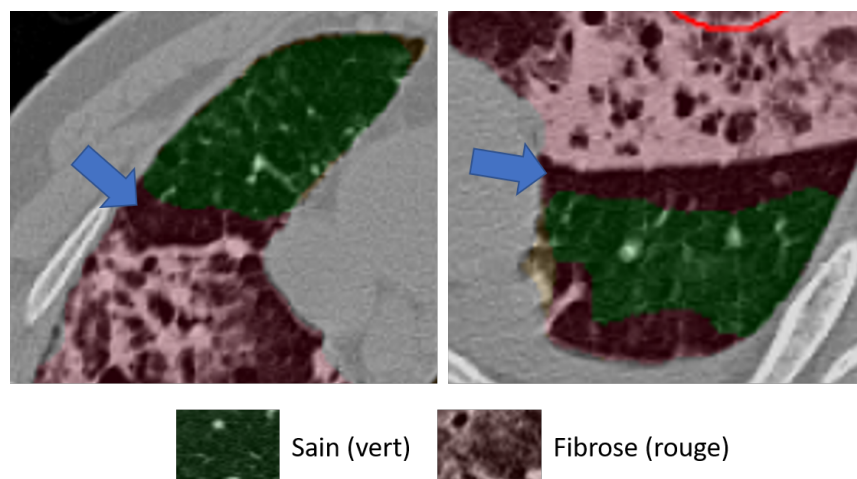


FIGURE 6.4. Sur les régions d'intérêt extraites de la segmentation présentée à la figure 5.3, les frontières entre les tissus sains et la fibrose incluent des bordures de tissus sains (flèches bleues) dans l'assignation de fibrose.

Les CNN entraînés (LungNet, UNet-Progressive et UNet-Progressive-Reg) peuvent segmenter adéquatement les tissus de fibrose et sains selon 3 classes (Sain, Verre dépoli, Fibrose). L'essentiel de la compréhension de l'image peut être effectuée par chacun des réseaux. Dans les cas où les motifs sont moins diversifiés et plus subtils, les segmentations sont très similaires. Un tel exemple est présenté à la figure 5.5. Le poumon sain, qui contient plusieurs vaisseaux, est correctement segmenté par tous les réseaux, de même que toutes les régions de verre dépoli.

La principale constatation est qu'il semble y avoir une confusion entre le poumon sain et le verre dépoli avec le LungNet. Dans les cas plus complexes, les segmentations des différents réseaux peuvent varier considérablement. Malgré que toutes les segmentations semblent valides, il est possible de constater que du côté de LungNet, il y a très peu de tissus sains assignés et la segmentation est bruitée (figure 6.7).

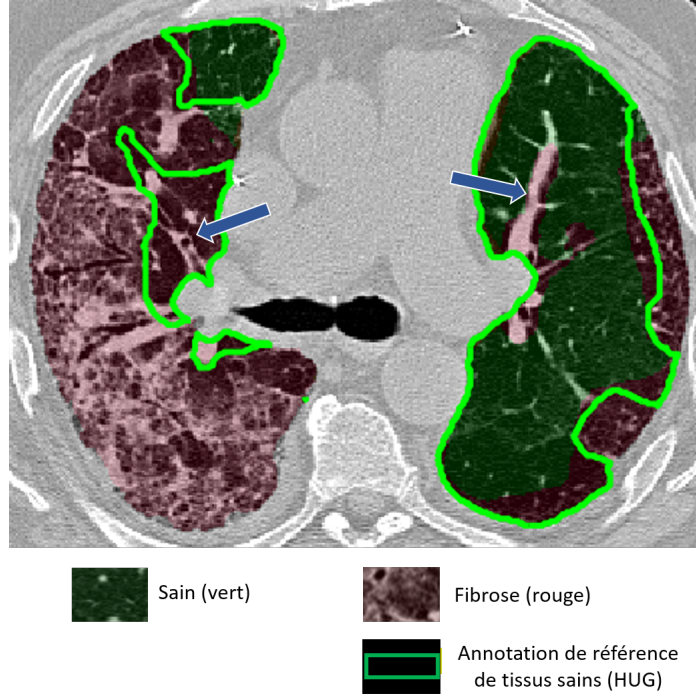


FIGURE 6.5. Des assignations erronées surviennent dans les régions près de l'arbre trachéobronchique (flèches bleues). Alors que ces régions sont saines, les motifs présents sont assignés comme étant des tissus de fibrose. Les contours verts correspondent aux réalités de terrain de tissus sains.

Les matrices de confusion générées à partir des différents ensembles de validation (figure 5.11) informent sur les différents types d'erreurs d'assignations faites par les différents réseaux. D'abord, il est possible de constater que les 3 CNN assignent correctement les voxels par rapport aux annotations partielles. Les coefficients Dice suggèrent que les réalités de terrain partielles et les prédictions du réseau se superposent de façon importante et que les segmentations comportent peu de faux positifs et de faux négatifs (pour UNet-Progressive-Reg, $DSC_{S,Reg} = 0.98 \pm 0.02$, $DSC_{VD,Reg} = 0.90 \pm 0.06$ et $DSC_{F,Reg} = 0.91 \pm 0.03$). Parmi tous les réseaux, les fractions d'assignations

valides les plus faibles sont de 92.6 % pour les tissus sains, 87.2 % pour le verre dépoli et 89.5 % pour la fibrose. Dans le même ordre d'idées, les justesses d'assignations sont de $A_{\text{Reg}} = 0.93 \pm 0.04$, $A_{\text{Prog.}} = 0.88 \pm 0.07$ et $A_{\text{LungNet}} = 0.89 \pm 0.05$.

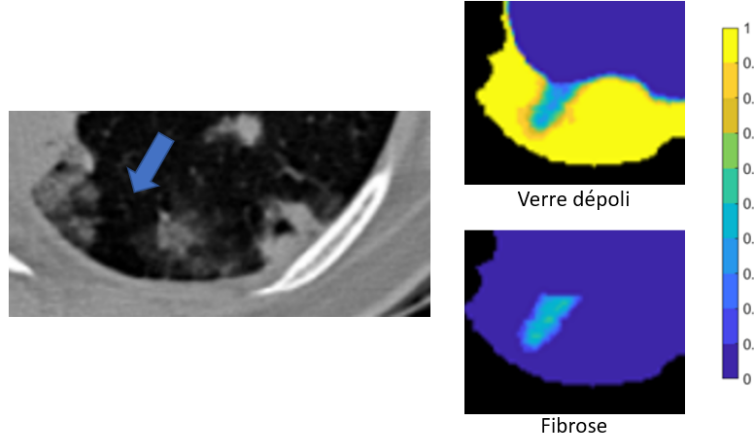


FIGURE 6.6. Sur la région d'intérêt extraite de la coupe présentée à la figure 5.2 (à gauche), il est possible de voir un cas (flèche bleue) où il y a une confusion entre le verre dépoli (en haut à droite) et la fibrose (en bas à droite) dans la prédiction. Sur les cartes softmax, on peut constater que les activations de verre dépoli et de fibrose sont similaires.

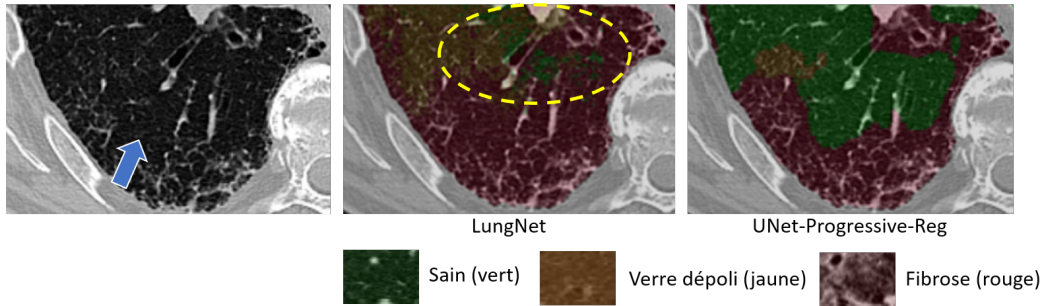


FIGURE 6.7. Sur la région d'intérêt extraite de la coupe présentée à la figure 5.6 (à gauche), il est possible de voir un exemple où LungNet assigne peu de tissus sains, comparativement à ce qui se trouve sur la région (flèche bleue). Aussi, il est possible de voir que la segmentation est bruitée (cercle jaune).

Tel que suggéré par les matrices de confusion, les assignations erronées entre la fibrose et les tissus sains sont très rares. Il est important de noter que les réseaux UNet-Progressive et UNet-Progressive-Reg segmentent les voxels sains avec très peu d'erreurs d'assignations (1.0 % et 0.7 % d'assignations erronées respectivement). Cette propriété est particulièrement intéressante dans le projet actuel, où il est tout aussi important d'assigner les tissus sains de façon performante.

La classe qui introduit le plus de difficultés dans la segmentation est le verre dépoli. Dans tous les cas, les voxels de verre dépoli sont bien identifiés et il y a relativement peu de faux négatifs, ce qui se traduit par des recouvrements élevés ($Re_{VD,Reg} = 0.92 \pm 0.03$, $Re_{VD,Prog.} = 0.87 \pm 0.11$ et $Re_{VD,LungNet} = 0.84 \pm 0.12$). Par contre, une fraction plus importante des autres tissus sont confondus avec le verre dépoli. Dans tous les cas, les faux positifs sont majoritairement causés les voxels de fibrose, mais dans le cas de LungNet, 6.3 % des voxels sains sont également des faux positifs par rapport au verre dépoli.

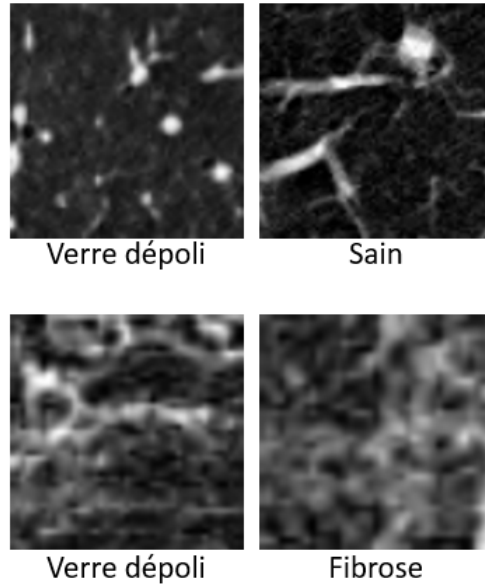


FIGURE 6.8. Les voxels où le verre dépoli est léger peuvent être confondus avec du tissu sain (en haut), tandis qu'il peut y avoir une confusion avec la fibrose lorsque celui-ci est intense (en bas). Des situations comme celles-ci sont susceptibles d'introduire des faux positifs par rapport au verre dépoli.

Ces confusions entre les autres tissus et le verre dépoli surviennent principalement à cause de la définition de la classe Verre dépoli. Celle-ci inclut tous les épaissements plus ou moins uniformes des tissus pulmonaires, pouvant être très légers (similaires à des tissus sains) ou plus intenses (similaires à de la fibrose). La figure 6.8 présente des exemples de ces deux situations. Cela se traduit par des précisions plus faibles, comparativement aux 2 autres classes ($Pr_{VD,Reg} = 0.87 \pm 0.09$, $Pr_{VD,Prog.} = 0.83 \pm 0.13$ et $Pr_{VD,LungNet} = 0.81 \pm 0.10$).

De plus, le UNet-Progressive-Reg a assigné le verre dépoli à chacun des volets avec des recouvrements similaires. En regardant les matrices de confusion de différents volets (figure 5.10), il est possible de voir que celui-ci a effectué peu d’erreurs d’assignments avec le verre dépoli. Par exemple, pour le volet (3), les fractions de tissus de verre dépoli mal assignés sont de 10.4 %, 36.4 % et 30.9 % pour UNet-Progressive-Reg, UNet-Progressive et LungNet respectivement. Les différentes métriques et les cartes de segmentation suggèrent que l’utilisation de l’architecture UNet-Progressive permet d’atteindre des performances comparables à LungNet pour l’assignation du verre dépoli et de la fibrose, tout en offrant des performances particulièrement intéressantes au niveau des tissus sains. Cette amélioration au niveau des tissus sains est également perceptible sur les cartes de segmentation hors des régions annotées. De façon générale, les cartes de segmentation semblent également moins bruitées. Ces améliorations peuvent notamment être dues à l’utilisation de petits facteurs de dilatation dans UNet-Progressive, qui permettent d’éviter un problème de maillage et de considérer adéquatement l’information du voisinage.

Aussi, le UNet-Progressive a un nombre élevé de paramètres à déterminer durant l’entraînement (2.0×10^7 paramètres), alors que LungNet en a considérablement moins (1.3×10^5 paramètres). Comme la quantité de données disponibles est limitée, l’ajout des éléments de régularisation (augmentation de données, dégradation des poids, *dropout*) a été bénéfique. Aussi, l’utilisation d’une fonction coût semi-supervisée a permis de considérer la totalité des voxels dans l’entraînement, plutôt que de se limiter aux annotations partielles. Le UNet-Progressive-Reg peut segmenter les tissus sains avec des performances similaires au UNet-Progressive, tout en s’améliorant au niveau des tissus de

verre dépoli ainsi qu’au niveau des tissus de fibrose ($DSC_{S,\text{Reg}} = 0.98 \pm 0.02$, $DSC_{VD,\text{Reg}} = 0.90 \pm 0.06$, $DSC_{F,\text{Reg}} = 0.91 \pm 0.03$).

Avec le UNet-Progressive-Reg, la majorité des erreurs d’assignations ont lieu entre le verre dépoli et la fibrose. Dans cette optique, l’utilisation du UNet-Progressive-Reg dans un contexte où ces deux classes sont combinées pourrait être particulièrement intéressante dans certaines applications en radio-oncologie. Dans ce scénario, deux classes seraient considérées, soit les tissus sains et les tissus pouvant être associés à de la fibrose pulmonaire (RFP). L’information fournie par le CNN serait moins spécifique, puisque la segmentation serait effectuée pour 2 classes, mais les erreurs d’assignation seraient grandement réduites, ce qui peut être plus bénéfique qu’une segmentation à 3 classes dans certaines applications de radio-oncologie. Par exemple, dans une application où une segmentation automatique est générée pour faciliter l’inspection visuelle des radio-oncologues, il peut être pertinent d’identifier plus exactement les zones à risque.

Dans cet ordre d’idées, les 3 patients de test ont été complètement annotés selon ces 2 classes afin d’évaluer les segmentations sur les poumons entiers plutôt que sur les régions partielles sélectionnées par les radiologistes. Globalement, le UNet-Progressive-Reg a pu assigner correctement la majorité des voxels de chacune des classes ($A = 0.88 \pm 0.06$) et les tissus de fibrose pulmonaire ont pu être segmentés avec un DSC de 0.89 ± 0.08 . Les patients 1 et 2 ont pu être segmentés correctement par rapport aux annotations complètes (figures 5.13 et 5.14), avec des coefficients Dice pour les classes Sain et RFP donnés par $DSC_S = 0.77$ et $DSC_{\text{RFP}} = 0.89$ pour le patient 1, puis $DSC_S = 0.97$ et $DSC_{\text{RFP}} = 0.81$ pour le patient 2. Pour le patient 3, dont la grande majorité des tissus correspond à de la fibrose sévère, des coefficients Dice de $DSC_S = 0.39$ et $DSC_{\text{RFP}} = 0.77$ ont été obtenus. L’échec de la segmentation des tissus sains dans le patient 3 est d’abord due au fait que ce patient correspond à un cas extrême de fibrose pulmonaire et que très peu de voxels sains sont présents. Tel qu’indiqué dans le tableau 4.3, 2.5×10^4 voxels sont annotés comme étant sains, tandis que 3.5×10^5 voxels correspondent à de la fibrose. De plus, les régions identifiées comme étant saines sont petites et le voisinage correspond à de la fibrose sévère (figure 5.15).

De façon générale, les métriques évaluées sur les tissus sains des patients de test sont plus faibles que sur les ensembles partiellement annotés. Cela peut être expliqué par une différence dans la définition des tissus sains utilisée dans les deux cas. Pour ce projet, l'utilisation des annotations partielles a été essentielle pour effectuer l'entraînement avec une plus grande quantité de données et des motifs plus diversifiés qu'en annotant complètement quelques patients. Toutefois, cela implique que, dans les annotations de test, plusieurs types de tissus peu représentés dans les annotations partielles sont inclus, tel que l'arbre trachéobronchique. À moins de procéder à une segmentation manuelle ou automatique de l'arbre trachéobronchique avant la segmentation de la fibrose, celui-ci devrait être inclus dans les annotations saines. Les similarités entre certains motifs de fibrose et l'arbre trachéobronchique contribuent à ce que l'assignation de ces régions soit souvent erronée. Néanmoins, des difficultés de ce genre sont également rencontrées lors des segmentations manuelles (figure 6.9).

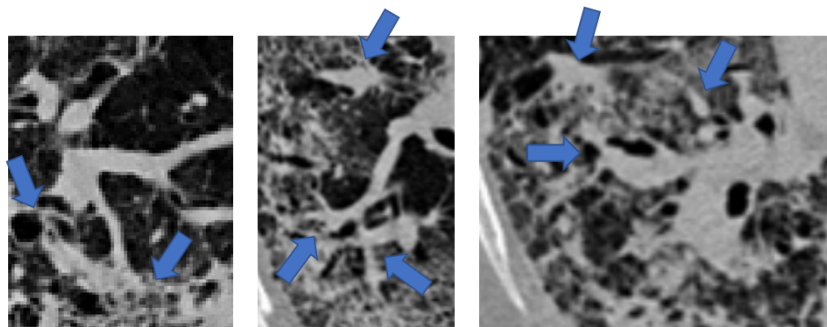


FIGURE 6.9. Trois régions d'intérêt qui illustrent des situations où la délimitation de l'arbre trachéobronchique et de la fibrose est ambiguë.

Aussi, dans l'ensemble partiellement annoté, les délinéations saines sont sélectionnées sur peu de patients dont les tissus sont majoritairement en bonne santé et couvrent la majeure partie des poumons. Sur les annotations complètes, plusieurs scénarios absents des données d'entraînement sont présents. Entre autres, les tissus pulmonaires identifiés comme étant sains sont plus diversifiés. Par exemple, des tissus sains peuvent être entourés de fibrose sévère, mais comme ces situations sont absentes de l'entraînement, le contexte peut induire le CNN en erreur, tel que chez le patient 3.

En résumé, les deux premiers patients de test ont pu être segmentés correctement, comme suggéré par les métriques évaluées et les cartes de segmentation. Les performances sur le patient 3 illustrent une limitation au niveau de la segmentation des tissus sains sur des cas très sévères de fibrose pulmonaire. Toutefois, d'un point de vue des applications cliniques envisagées, les deux premiers patients représentent mieux des situations où l'algorithme de segmentation serait réellement utilisé, soit des cas de fibrose pulmonaire plus légers et pouvant contenir d'autres pathologies. Néanmoins, une solution pour améliorer les performances à proximité de l'arbre trachéobronchique et dans les régions entourées de fibrose sévère serait de produire de nouvelles annotations ou de bonifier les annotations existantes. Au cours de ce projet, l'annotation d'une base de données en radio-oncologie de patients de fibrose a été amorcée et pourrait servir de point de départ. L'objectif serait d'inclure une plus grande diversité de tissus dans les tissus sains utilisés dans l'entraînement.

Pour la suite du projet, plusieurs applications peuvent être considérées. Dans un premier temps, un réseau entraîné pourrait être utilisé pour segmenter automatiquement un patient entier dans le but de faciliter l'appréciation visuelle des images par les radio-oncologues. Le réseau pourrait être utilisé pour segmenter les tissus selon 2 ou 3 classes, avec ou sans seuil softmax, selon les besoins. Étant donné les effets néfastes qui peuvent survenir suite à l'irradiation de poumons atteints de fibrose pulmonaire, son utilisation peut aider à faire un dépistage des patients avant d'effectuer une planification de traitement. Une autre application envisageable est le suivi de la progression de la fibrose chez un patient au cours d'un traitement de radiothérapie. Dans le même ordre d'idées, le suivi peut être fait post-traitement, pour évaluer si la radiothérapie a causé une apparition de fibrose. À plus long terme, il serait possible d'intégrer l'information fournie par l'algorithme d'assignation de tissus dans la planification de traitements. L'information pourrait être utilisée, par exemple, pour éviter les régions atteintes de fibrose ou, à l'inverse, pour épargner les régions plus fonctionnelles des poumons. De plus, il serait possible de considérer la segmentation automatique d'un plus grand nombre de tissus pathologiques, par exemple l'emphysème, dans le but d'acquérir davantage d'information pertinente cliniquement sur la présence de maladies pulmonaires.

Pour améliorer davantage les performances ou ouvrir la porte à d'autres applications, un premier élément à considérer est l'annotation de patients provenant d'une base de données de radio-oncologie. Intégrer des données de radio-oncologie permettrait d'abord d'augmenter la diversité et la quantité des données disponibles. Comme le UNet-Progressive-Reg est relativement complexe, accroître le nombre de coupes durant l'entraînement pourrait améliorer directement les performances. Aussi, accroître la quantité de données peut permettre d'apporter des modifications basées sur une augmentation du nombre de paramètres à l'architecture. Des modifications possibles incluent de considérer 64 cartes d'activations initiales plutôt que 32, d'ajouter des modules ou d'étendre ceux-ci. Avec l'élaboration d'une base de données où toutes les coupes d'un patient sont disponibles, le réseau pourrait être adapté pour analyser des volumes plutôt que des coupes individuelles. Un autre élément qui pourrait être considéré est l'utilisation de modules d'inception. Un module d'inception consiste, en résumé, à connecter la couche précédente du réseau à plusieurs opérations parallèles, suivi d'une concaténation des cartes. Par exemple, les différentes opérations peuvent faire varier la taille du noyau de convolution. Aussi, une optimisation du réseau pourrait être effectuée à l'aide d'un algorithme génétique. De plus, introduire un seuil aux activations softmax pourrait également être bénéfique. Les voxels dont l'activation softmax de prédiction est inférieure au seuil ne recevraient pas d'assignation dans les cartes de segmentation finales.

Lors de l'annotation de la base de données, il serait essentiel de définir de façon appropriée la classe des tissus sains pour qu'elle reflète bien les patients rencontrés en radio-oncologie. Pour étendre la segmentation au-delà de la fibrose pulmonaire, d'autres pathologies pulmonaires pourraient être considérées, telles que l'emphysème. Dans le même ordre d'idées, les rayons de miel et la réticulation pourraient être découplés de la classe Fibrose. Finalement, dans les données utilisées au cours de ce projet, les champs de vue utilisés durant la reconstruction varient grandement. Donc, les dimensions physiques des voxels et des caractéristiques fluctuent de façon importante entre les patients. En radio-oncologie, les voxels ont généralement les mêmes dimensions pour tous les patients, mais le champ de vue est considérablement plus élevé. Pour avoir une qualité d'image qui s'approche des données haute résolution, il pourrait être

pertinent d'effectuer une reconstruction des images de radio-oncologie avec un champ de vue plus petit, en plus des reconstructions utilisées actuellement.

Chapitre 7

CONCLUSION

Au cours de ce projet, des outils permettant de segmenter automatiquement la fibrose pulmonaire sur des images CT ont été développés. L'utilisation de réseaux de neurones complètement convolutifs permet de générer des prédictions denses, pour tous les voxels d'une image, en une seule inférence. Ces réseaux peuvent prendre en entrée une image CT complète et produire en sortie une carte d'activations softmax par type de tissu considéré. Cela permet d'assigner des tissus sains, de verre dépoli et de fibrose aux voxels pulmonaires de façon efficace. Le CNN UNet-Progressive-Reg a pu segmenter les données provenant de la base de données d'images CT haute résolution avec des coefficients Dice supérieurs à 0.90 pour chacune des classes considérées. Ce CNN, basé sur les convolutions dilatées progressivement et l'architecture U-Net, inclut plus d'une trentaine de couches de convolutions, organisées en un encodeur, un pont et un décodeur. Plusieurs éléments destinés à améliorer la capacité de généralisation du réseau sont intégrés, tels que les connexions raccourcis, les unités linéaires rectifiées et la normalisation des activations. Plusieurs éléments de régularisation sont utilisés durant la phase d'entraînement, comme la dégradation des poids, l'abandon de neurones et l'augmentation de données. Tous les voxels, annotés ou non, ont pu être considérés lors l'entraînement, par l'utilisation d'une fonction coût semi-supervisée. L'objectif du projet, qui était de développer et d'implémenter des modèles permettant de segmenter automatiquement des tissus associés à de la fibrose pulmonaire, a donc été atteint. Les performances obtenues avec les modèles implémentés suggèrent que ceux-ci pourraient être intéressants en tant que base dans une application d'aide aux médecins, où les patients pourraient être pré-segmentés par un réseau dans le but de faciliter l'interprétation visuelle par les médecins.

La segmentation automatique de la fibrose pulmonaire ouvre la porte à de nombreuses applications futures en radio-oncologie. Dans un premier temps, avant la planification de traitement, une segmentation du réseau peut être utilisée dans le cadre d'une aide aux radio-oncologues face au dépistage de la

fibrose. Étant donné les effets secondaires néfastes pouvant survenir suite à une irradiation de tissus de fibrose, l'information procurée par ces segmentations peut mener à une modification des méthodes de traitement considérées. La segmentation de la fibrose peut également être utilisée dans le suivi de l'évolution de la fibrose au cours d'un traitement de radiothérapie, ainsi que post-traitement. À plus long terme, l'information sur la présence de fibrose pourrait être intégrée dans la planification des traitements pour, par exemple, éviter les régions atteintes de fibrose ou, à l'inverse, pour épargner les régions plus fonctionnelles du poumon. Plusieurs options de recherches futures ont été abordées dans le chapitre 6. D'abord, il serait essentiel de produire de nouvelles données avec plus d'annotations que ce qui est disponible actuellement. Les annotations d'une base de données de radio-oncologie amorcées au cours de ce projet pourraient servir de point de départ. Cela permettrait d'abord d'améliorer les performances à proximité de l'arbre trachéobronchique et sur les tissus sains plus subtils, qui sont actuellement limitées par l'absence d'annotations dans ces zones. D'autre part, cela ouvrirait la porte à une augmentation du nombre de types de tissus considérés et des modifications pourraient être apportées au UNet-Progressive-Reg. Avec davantage de données disponibles, une augmentation du nombre de paramètres du réseau pourrait être considérée. En considérant la totalité des coupes pulmonaires, l'architecture pourrait être étendue pour analyser et segmenter des volumes pulmonaires plutôt que des coupes axiales. Avec les données nécessaires, il serait même envisageable d'étendre l'outil actuel au-delà de la fibrose pulmonaire et de développer une plateforme de dépistage de pathologies pulmonaires.

Bibliographie

- [1] Société canadienne du Cancer. Vue d'ensemble des statistiques sur le cancer. "<http://www.cancer.ca/fr-ca/cancer-information/cancer-101/cancer-statistics-at-a-glance/?region=qc>". Consulté en 2018.
- [2] Kong F.-M., Ten Haken R. K., Schipper M. J., Sullivan M. A. et al. High-dose radiation improved local tumor control and overall survival in patients with inoperable/unresectable non-small-cell lung cancer : Long-term results of a radiation dose escalation study. *International Journal of Radiation Oncology, Biology, Physics*, 63(2) :324–333, 2005.
- [3] Ozawa Y., Abe T., Omae M., Matsui T. et al. Impact of preexisting interstitial lung disease on acute, extensive radiation pneumonitis : Retrospective analysis of patients with lung cancer. *PLOS One*, 10(10) : 1–12, 2015.
- [4] Bahig H., Filion E., Vu T., Chalaoui J. et al. Severe radiation pneumonitis after lung stereotactic ablative radiation therapy in patients with interstitial lung disease. *Practical Radiation Oncology*, 6(5) :367–374, 2016.
- [5] Bourke S. Interstitial lung disease : progress and problems. *Postgraduate Medical Journal*, 82(970) :494–499, 2006.
- [6] Depeursinge A., Vargas A., Platon A., Geissbuhler A. et al. Building a reference multimedia database for interstitial lung diseases. *Computerized Medical Imaging and Graphics*, 36(3) :227–238, 2012.
- [7] du Bois R. M. An earlier and more confident diagnosis of idiopathic pulmonary fibrosis. *European Respiratory Review*, 21(124) :141–146, 2012.
- [8] Fujimoto H., Kobayashi T. et Azuma A. Idiopathic Pulmonary Fibrosis : Treatment and Prognosis. *Clinical Medicine Insights : Circulatory, Respiratory and Pulmonary Medicine*, 9(Suppl 1) :179–185, 2015.
- [9] Gilman M. J., Laurens R. G., Somogyi J. W. et Honig E. G. CT attenuation values of lung density in sarcoidosis. *Journal of Computer Assisted Tomography*, 7(3) :407–410, 1983.

- [10] Best A. C., Lynch A. M., Bozic C. M., Miller D. et al. Quantitative CT indexes in idiopathic pulmonary fibrosis : Relationship with physiologic impairment. *Radiology*, 228(2) :407–414, 2003.
- [11] Ohkubo H., Nakagawa H. et Niimi A. Computer-based quantitative computed tomography image analysis in idiopathic pulmonary fibrosis : A mini review. *Respiratory Investigation*, 56(1) :5–13, 2018.
- [12] Wang Z., Gu S., Leader J. K., Kundu S. et al. Optimal threshold in CT quantification of emphysema. *European Radiology*, 23(4) :975–984, 2013.
- [13] Müller N. L., Staples C. A., Miller R. R. et Abboud R. T. "Density mask" : An objective method to quantitate emphysema using computed tomography. *CHEST*, 94(4) :782–787, 1988.
- [14] Shin K. E., Chung M. J., Jung M. P., Choe B. K. et Le K. S. Quantitative computed tomographic indexes in diffuse interstitial lung disease : correlation with physiologic tests and computed tomography visual scores. *Journal of Computer Assisted Tomography*, 35(2) :266–271, 2011.
- [15] Ohkubo H., Kanemitsu Y., Uemura T., Takakuwa O. et al. Normal Lung Quantification in Usual Interstitial Pneumonia Pattern : The Impact of Threshold-based Volumetric CT Analysis for the Staging of Idiopathic Pulmonary Fibrosis. *PLOS One*, 11(3) :e0152505, 2016.
- [16] Nakagawa H., Nagatani Y., Takahashi M., Ogawa E., Tho N. V., Ryujin Y., Nagao T. et Nakano Y. Quantitative CT analysis of honeycombing area in idiopathic pulmonary fibrosis : Correlations with pulmonary function tests. *European Journal of Radiology*, 85(1) :125–130, 2016.
- [17] Kim N., Seo J. B., Lee Y., Lee J. G. et al. Development of an automatic classification system for differentiation of obstructive lung disease using HRCT. *Journal of Digital Imaging*, 22(2) :136–148, 2009.
- [18] Xu Y., van Beek E. J., Hwanjo Y., Guo J. et al. Computer-aided classification of interstitial lung diseases via MDCT : 3D adaptive multiple feature method (3D AMFM). *Academic Radiology*, 13(8) :969–978, 2006.

- [19] Depeursinge A., Chin A. S., Leung A. N., Terrone D. et al. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography. *Investigative Radiology*, 50(4) :261–267, 2015.
- [20] Bartholmai B. J., Raghunath S., Karwoski R. A., Moua T., Rajagopalan S., Maldonado F., Decker P. A. et Robb R. A. Quantitative computed tomography imaging of interstitial lung diseases. *Journal of Thoracic Imaging*, 28(5) :298–307, 2013.
- [21] van Ginneken B., Katsuragawa S., ter Haar Romeny B. M., Doi K. et Viergever M. A. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Transactions on Medical Imaging*, 21(2) : 139–149, 2002.
- [22] Peng S. H., Kim D. H., Lee S. L. et Lim M. K. Texture feature extraction based on a uniformity estimation method for local brightness and structure in chest CT images. *Computers in Biology and Medicine*, 40(11-12) :931–942, 2010.
- [23] Liang T. K., Tanaka T., Nakamura H. et Ishizaka A. A neural network based computer-aided diagnosis of emphysema using CT lung images. Dans *SICE Annual Conference 2007*, 2007.
- [24] Dudhane A., Shingadkar G., Sanghavi P., Jankharia B. et Talbar S. Interstitial lung disease classification using feed forward neural networks. Dans *International Conference on Communication and Signal Processing 2016 (ICCASP 2016)*, 2016.
- [25] Kim K. G., Goo J. M., Kim J. H., Lee H. J. et al. Computer-aided diagnosis of localized ground-glass opacity in the lung at CT : initial experience. *Radiology*, 237(2) :657–661, 2005.
- [26] Anthimopoulos M., Christodoulidis S., Ebner L., Christe A. et Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging*, 35 (5) :1207–1216, 2016.

- [27] Shin H. C., Roth H. R., Gao M., Lu L. et al. Deep Convolutional Neural Networks for Computer-Aided Detection : CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5) :1285–1298, 2016.
- [28] Anthimopoulos M., Christodoulidis S., Ebner L., Geiser T., Christe A. et Mougiakakou S. Semantic Segmentation of Pathological Lung Tissue With Dilated Fully Convolutional Networks. *IEEE Journal of Biomedical and Health Informatics*, 23(2) :714–722, 2019.
- [29] Johns H. E. et Cunningham J. R. *The physics of radiology*. Charles C. Thomas, 1983.
- [30] Andreo P., Burns D. T., Nahum A. E., Seuntjens J. et Attix F. H. *Fundamentals of Ionizing Radiation Dosimetry*. Wiley, 2017.
- [31] Attix F. H. *Introduction to Radiological Physics and Radiation Dosimetry*. Wiley, 1986.
- [32] Bushberg J. T. et Boone J. M. *The Essential Physics of Medical Imaging*. Wolters Kluwer Health, 2011.
- [33] Podgorsak E. B. *Radiation Physics for Medical Physicists*. Springer Berlin Heidelberg, 2010.
- [34] Geyer L. L., Schoepf U. J., Meinel F. G., Nance J. W. et al. State of the Art : Iterative CT reconstruction techniques. *Radiology*, 276(2) :339–357, 2015.
- [35] Gotway M. B., Reddy G. P., Webb W. R., Elicker B. M. et Leung J. W. High-resolution CT of the lung : patterns of disease and differential diagnoses. *Radiologic Clinics of North America*, 43(3) :513–542, 2005.
- [36] McCulloch W. S. et Pitts W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4) :115–133, 1943.
- [37] Rosenblatt F. The Perceptron : A Probabilistic Model for Information Storage and Organization in The Brain. *Psychological Review*, 65(6) :65–386, 1958.
- [38] Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4) :303–314, 1989.

- [39] Hornik K., Stinchcombe M. et White H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5) :359–366, 1989.
- [40] Sonoda S. et Murata N. Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, 43(2) :233–268, 2017.
- [41] Goodfellow I., Bengio Y. et Courville A. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [42] Kingma D. P. et Ba J. Adam : A Method for Stochastic Optimization. *arXiv preprint arXiv :1412.6980*, 2014.
- [43] Nielsen M. A. *Neural networks and deep learning*. Determination Press, 2015.
- [44] Nair V. et Hinton G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. Dans *27th International Conference on International Conference on Machine Learning (ICML)*, pages 807–814. Omnipress, 2010.
- [45] Glorot X., Bordes A. et Bengio Y. Deep Sparse Rectifier Neural Networks. Dans *Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15, pages 315–323. PMLR, 2011.
- [46] Zeiler M. D., Ranzato M., Monga R., Mao M., Yang K., Le Q. V., Nguyen P., Senior A., Vanhoucke V., Dean J. et Hinton G. E. On rectified linear units for speech processing. Dans *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3517–3521, 2013.
- [47] LeCun Y., Bengio Y. et Hinton G. Deep learning. *Nature*, 521(7553) :436–444, 2015.
- [48] Maas A. L., Hannun A. Y. et Ng A. Y. Rectifier nonlinearities improve neural network acoustic models. Dans *ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.
- [49] He K., Zhang X., Ren S. et Sun J. Delving Deep into Rectifiers : Surpassing Human-Level Performance on ImageNet Classification. Dans *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034. IEEE Computer Society, 2015.
- [50] Perez L. et Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv e-prints arXiv :1712.04621*, 2017.

- [51] LeCun Y., Cortes C. et Burges C. The MNIST Database of Handwritten Digits. "<http://yann.lecun.com/exdb/mnist/>". Consulté en 2019.
- [52] Simard P., Steinkraus D. et Platt J. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. Dans *Seventh International Conference on Document Analysis and Recognition*, 2003.
- [53] Bishop C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [54] Stanford University. CS231n : Convolutional Neural Networks for Visual Recognition. "<http://cs231n.github.io/>". Consulté en 2019.
- [55] Srivastava N., Hinton G., Krizhevsky A., Sutskever I. et Salakhutdinov R. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15 :1929–1958, 2014.
- [56] Bengio Y., Simard P. et Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2) : 157–166, 1994.
- [57] Hochreiter S., Bengio Y. et Frasconi P. Gradient Flow in Recurrent Nets : the Difficulty of Learning Long-Term Dependencies. Dans *Field Guide to Dynamical Recurrent Networks*. IEEE Press, 2001.
- [58] Yu F. et Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions. Dans *4th International Conference on Learning Representations (ICLR)*, 2016.
- [59] Holschneider M., Kronland-Martinet R., Morlet J. et Tchamitchian P. A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform. Dans *Wavelets*, pages 286–297. Springer Berlin Heidelberg, 1990.
- [60] Dutilleul P. An Implementation of the "algorithme à trous" to Compute the Wavelet Transform. Dans *Wavelets*, pages 298–304. Springer Berlin Heidelberg, 1990.
- [61] Yu F., Koltun V. et Funkhouser T. Dilated Residual Networks. Dans *Computer Vision and Pattern Recognition (CVPR)*, 2017.

- [62] Wang P., Chen P., Yuan Y., Liu D., Huang Z., Hou X. et Cottrell G. Understanding Convolution for Semantic Segmentation. Dans *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1451–1460, 2018.
- [63] Romera E., Álvarez J. M., Bergasa L. M. et Arroyo R. Efficient ConvNet for real-time semantic segmentation. Dans *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1789–1794, 2017.
- [64] Paszke A., Chaurasia A., Kim S. et Culurciello E. ENet : A Deep Neural Network Architecture for Real-Time Semantic Segmentation. *CoRR*, abs/1606.02147, 2016.
- [65] Dolz J., Xu X., Rony J., Yuan J. and Liu Y., Granger E., Desrosiers C., Zhang X., Ben Ayed I. et Lu H. Multiregion segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks. *Medical Physics*, 45(12) :5482–5493, 2018.
- [66] Lopez M. M., Ventura J., Bakas S., Kuijf H. et al. Dilated Convolutions for Brain Tumor Segmentation in MRI Scans. Dans *Brainlesion : Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes)*, pages 253–262. Springer International Publishing, 2018.
- [67] Mehta S., Rastegari M., Caspi A., Shapiro L. et Hajishirzi H. ESPNet : Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation. Dans *The European Conference on Computer Vision (ECCV)*, September 2018.
- [68] Yamashita T., Furukawa H. et Fujiyoshi H. Multiple Skip Connections of Dilated Convolution Network for Semantic Segmentation. Dans *25th IEEE International Conference on Image Processing (ICIP)*, pages 1593–1597, 2018.
- [69] Hamaguchi R., Fujita A., Nemoto K., Imaizumi T. et Hikosaka S. Effective Use of Dilated Convolutions for Segmenting Small Object Instances in Remote Sensing Imagery. Dans *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1442–1450, 2018.

- [70] Chen L., Papandreou G., Kokkinos I., Murphy K. et Yuille A. L. DeepLab : Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4) :834–848, 2018.
- [71] Ioffe S. et Szegedy C. Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift. Dans *32nd International Conference on Machine Learning (ICML)*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, 2015.
- [72] Kohler J., Daneshmand H., Lucchi A., Hofmann T. et al. Exponential convergence rates for Batch Normalization : The power of length-direction decoupling in non-convex optimization. Dans *Proceedings of Machine Learning Research (PMLR)*, volume 89, pages 806–815, 2019.
- [73] Santurkar S., Tsipras D., Ilyas A. et Madry A. How Does Batch Normalization Help Optimization? Dans *Advances in Neural Information Processing Systems 31 (NIPS)*, pages 2483–2493. Curran Associates, Inc., 2018.
- [74] Isola P., Zhu J., Zhou T. et Efros A. A. Image-to-Image Translation with Conditional Adversarial Networks. Dans *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
- [75] Ulyanov D., Vedaldi A. et Lempitsky V. S. Improved Texture Networks : Maximizing Quality and Diversity in Feed-Forward Stylization and Texture Synthesis. Dans *2017 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, pages 4105–4113. IEEE Computer Society, 2017.
- [76] Wu Y. et He K. Group Normalization. Dans *15th European Conference on Computer Vision (ECCV)*, volume 11217, pages 3–19. Springer, 2018.
- [77] He K., Zhang X., Ren S. et Sun J. Deep Residual Learning for Image Recognition. Dans *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [78] Drozdal M., Vorontsov E., Chartrand G., Kadoury S. et Pal C. The Importance of Skip Connections in Biomedical Image Segmentation. Dans *Deep Learning and Data Labeling for Medical Applications*, pages 179–187. Springer International Publishing, 2016.

- [79] Dumoulin V. et Visin F. A guide to convolution arithmetic for deep learning. *ArXiv e-prints*, 2016.
- [80] Long J., Shelhamer E. et Darrell T. Fully convolutional networks for semantic segmentation. Dans *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.
- [81] Ronneberger O., Fischer P. et Brox T. U-Net : Convolutional Networks for Biomedical Image Segmentation. Dans *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer International Publishing, 2015.
- [82] Badrinarayanan V., Kendall A. et Cipolla R. Segnet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12) :2481–2495, 2017.
- [83] Christ P. F., Elshaer M. E. A., Ettlinger F., Tatavarty S. et al. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. Dans *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 415–423. Springer International Publishing, 2016.
- [84] Dong H., Yang G., Liu F., Mo Y. et Guo Y. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. Dans *Medical Image Understanding and Analysis (MIUA)*, pages 506–517. Springer International Publishing, 2017.
- [85] Korsuk S., Josien P. W. P., Hao C., Xiaojuan Q. et al. Gland segmentation in colon histology images : The glas challenge contest. *Medical Image Analysis*, 35 :489–502, 2017.
- [86] Zotti C., Luo Z., Humbert O., Lalande A. et Jodoin P.-M. GridNet with Automatic Shape Prior Registration for Automatic Mri Cardiac Segmentation. Dans *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges (STACOM)*, pages 73–81, 2018.
- [87] Dice L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3) :297–302, 1945.

- [88] Taha A. A. et Hanbury A. Metrics for evaluating 3D medical image segmentation : analysis, selection, and tool. *BMC Medical Imaging*, 15 : 29, 2015.
- [89] Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z. et al. TensorFlow : Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>.
- [90] Glorot X. et Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Dans *Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 249–256, 2010.
- [91] Paszke A., Gross S., Chintala S., Chanan G. et al. Automatic differentiation in pytorch. 2017.
- [92] Tompson J., Goroshin R., Jain A., LeCun Y. et Bregler C. Efficient object localization using Convolutional Networks. Dans *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 648–656, 2015.
- [93] Chapelle O. et Zien A. Semi-supervised classification by low density separation. Dans *Tenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 57–64, 2005.
- [94] Dong-Hyun L. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. Dans *Workshop on Challenges in Representation Learning (ICML)*, volume 3, 2013.
- [95] Grandvalet Y. et Bengio Y. Semi-supervised learning by entropy minimization. Dans *Advances in Neural Information Processing Systems 17 (NIPS)*, pages 529–536. 2005.